The computational face for facial emotion analysis

Computer based emotion analysis from the face

Ahmad Al-DAHOUD

Submitted for the Degree of Doctor of Philosophy

Faculty of Engineering and Informatics

School of Media, Design and Technology

University of Bradford

2018

Abstract

Facial expressions are considered to be the most revealing way of understanding the human psychological state during face-to-face communication. It is believed that a more natural interaction between humans and machines can be undertaken through the detailed understanding of the different facial expressions which imitate the manner by which humans communicate with each other.

In this research, we study the different aspects of facial emotion detection, analysis and investigate possible hidden identity clues within the facial expressions. We study a deeper aspect of facial expressions whereby we try to identify gender and human identity - which can be considered as a form of emotional biometric - using only the dynamic characteristics of the smile expressions. Further, we present a statistical model for analysing the relationship between facial features and Duchenne (real) and non-Duchenne (posed) smiles. Thus, we identify that the expressions in the eyes contain discriminating features between Duchenne and non-Duchenne smiles.

Our results indicate that facial expressions can be identified through facial movement analysis models where we get an accuracy rate of 86% for classifying the six universal facial expressions and 94% for classifying the common 18 facial action units. Further, we successfully identify the gender using only the dynamic characteristics of the smile expression whereby we obtain an 86% classification rate. Likewise, we present a framework to study the possibility of using the smile as a biometric whereby we show that the human smile is unique and stable.

Dedications

To my Father and Mother,

To My wife and daughter,

To My Sisters,

To my friends.

Acknowledgments

All thankfulness to Allah, the most gracious and the most merciful, for providing me with the blessing to undertake research towards this PhD. I couldn't have done this without his mercy.

I would like to dedicate my deepest gratitude to my supervisor Prof. Hassan Ugail for his support, guidance, encouragement, constructive ideas and suggestions which helped me to complete this research and this thesis.

My special thanks go to my beloved parents for their support, love and encouragement throughout my life. I thank my father for his sponsorship and care; without him I wouldn't be here.

I would like to thank my wife for her patience and support through the most challenging period of my PhD research. To my daughter, the light in my life, and to my sisters for their cheerful support. For my aunts and uncles for love and prayer.

For my friends Ahmad Al-Salman and Mohammad Al-Nader for their positive influence in my life. To my colleagues Ranya Alzubaidi, Ali Elmahmudi, Zahra Sayed and Nosheen Hussain for their help, support and kindness.

I gratefully acknowledge Al-Zaytoonah University for their funding and support which made this PhD research possible in the first place.

Finally, for my Z28 car for keeping me enthusiastic, happy and joyful.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

HCI	Human computer interaction
PCI	Principle component analysis
AHCI	Affective human computer interaction
MFCCs	Mel Frequency Cepstral Coefficients
IEMOCAP	Emotional Dyadic Motion Capture database
GMMs	Gaussian Mixture Models
FACS	Facial Action Coding System
FAPs	facial animation parameters
AUs	Action Units'
AUDs	action unit descriptors
EMFACS	Emotional Facial Action Coding System
MPEG-4	Moving Picture Experts Group
FBA	Face and Body Animation
FDPs	Facial definition parameters
CGI	Computer generated imaginary
MUG	Multimedia Understanding Group
HMM	Hidden Markov model
AAM	Active Appearance Model
FCP	Facial Characteristic Points
SVM	Support vector machine
NN	Neural network
ROI	Region of interest
ASM	Active Shape Model
EMG	Electromyography
LBP	Local binary pattern
DCT	Discrete Cosine Transform
GDF	Geometrical Distance Feature
LDP	Local Directional Pattern
HOG	Histogram of Oriented Gradients
VCML	Covariance Matrix Logarithm

HOF	Histogram of Optical Flow
MBH	Motion Boundary Histogram
RDT-DWT	wo-level Dual Tree Discrete Wavelet Transform
BPNN	Back Propagation Neural Network
LGBPHS	local Gabor binary pattern histogram sequence
GMP	Gabor Magnitude Pictures
KNN	K-nearest neighbour
mRmR	Minimum redundancy maximum relevance
MV	Machine vision
FS	Floor sensor
WS	Wearable sensor
MV	Machine vision
GEI	Gait Energy Image
ACDA	Adaptive Component and Discriminant Analysis
FLD	Fisher linear discriminant
CSF	Circular symmetric filter
AAD	Average absolute deviation
NFL	Nearest feature line
FEB	Facial expression biometric
CK+	Cohn-Kanade’s extended facial expression
MVRE	Motion vector re-calculation engine
FNP	Facial normalisation parameter
CNN	Conventional neural network

List of Publications

Conference papers

1. Ahmad Al-dahoud and Hassan Ugail ***A Method for Location based Search for Enhancing Facial Feature Detection.*** Advances in Computational Intelligence Systems, September, 2016, pp 421-432.

2. Ahmad Al-dahoud and Hassan Ugail ***On Gender Identification Using the Smile Dynamics***," *2017 International Conference on Cyberworlds (CW)*, Chester, 2017, pp. 1-8. doi: 10.1109/CW.2017.26.

Journal paper

1. Hassan Ugail and Ahmad Al-dahoud ***Is gender encoded in the smile? A computational framework for the analysis of the smile driven dynamic face for gender recognition***, The Visual Computer, 2018/03/05, 2018.

Papers under review

1. Ahmad Al-dahoud and Hassan Ugail: ***Computational Analysis of Smile Weight Distribution across the Face for Accurate Distinction between Genuine and Posed Smiles.*** , submitted to *2018 International Conference on Cyberworlds*.

Papers being prepared

1. Ahmad Al-dahoud and Hassan Ugail ***A framework for detecting facial expression and action using flow of the face***.

2. Ahmad Al-dahoud and Hassan Ugail ***Facial expression biometric : is human identity encoded in the dynamic characteristic of the smile expression?***

1 Introduction

1.1 Emotions

"*Everyone knows what an emotion is, until asked to give a definition*"

Beverly Fehr and James Russell [1]

Emotions are very complex to define; there is no universal definition for them. The word 'emotion' is used to refer to feelings which are reflected on human physical appearance. The Oxford Dictionary defines it as: "*A strong feeling deriving from one's circumstances, mood, or relationships with others*". The Medical Dictionary goes into further detail with: "*A conscious mental reaction (as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioural changes in the body*". From a social science point of view, emotions are feelings and moods. Both feelings and moods represent emotions, but feelings are instant and short-lived, whereas moods are longer and leave a lasting expression. Feelings include experiences such as love, hate, anger, trust, joy, panic, fear and grief. The mood is a more general feeling such as happiness, sadness, frustration, contentment or anxiety.

The book, 'The Expression of the Emotions in Man and Animals', by Charles Darwin is considered to be one of the first pieces of research carried out on emotions; Darwin classifies emotions to be species-specific rather than culture-specific [2]. In 1969, Ekman and Friesen recognised the universality of emotions across different cultures; they considered six facial emotion expressions to be universal (happiness, sadness, anger, disgust, surprise and fear) [3].

Everyone experiences emotions, but each person experiences them in a different way. Furthermore, scientists do not all agree on the criteria for measuring or studying emotions because emotions are complex and have both physical and mental components. However, all researchers agree that emotions are composed of subjective feelings, physiological (body) responses and expressive behaviour. A subjective feeling is defined as an individual experience of emotion (personal experience); the subjective feeling is very hard to measure or describe. Each person's definition of emotion will not be the same, for example, two people experiencing anger will not experience or describe anger in the same way.

Physiological responses are specific changes in body functionality caused by the nervous system when facing a specific emotion. Physiological responses are simple to measure because scientists have developed tools to measure them; heart rate, sweating, blood pressure or realising adrenaline in blood are some of the many properties used to measure physiological responses. Research has shown that people have similar internal responses to the same emotion, regardless of their age, race or gender [4]. The internal response to stress is for the body to release the hormone adrenaline. This hormone helps prepare for the 'fight or flight' reaction, which means the body prepares to either run away or fight [5]. Research shows in general that the same emotions produce the same physical reaction.

Expressive behaviour refers to visible signs that an emotion is being experienced. These signs include: fainting, a flushed face, muscle stress, facial expressions and changes in the voice tone, breathing in a fast/slow manner or

other ‘body language’. These signs help to identify the emotional state and allow others to better interact with and understand a person’s emotion when dealing with a specific situation [6]. Based on subjective feelings, physiological responses and expressive behaviour there are many theories about how emotions are generated. By taking the facial expression as an example of generating emotion, sometimes humans express some facial expressions without being aware. These facial expressions are caused by a complex system of direct and indirect neural pathways to and from the motor cortex, brain stem and limbic system. Figure 1-1 shows the difference between a spontaneous smile and a voluntary smile, and the corresponding organs responsible for each of them. It also shows that different inner organs can trigger the same emotion [6] .

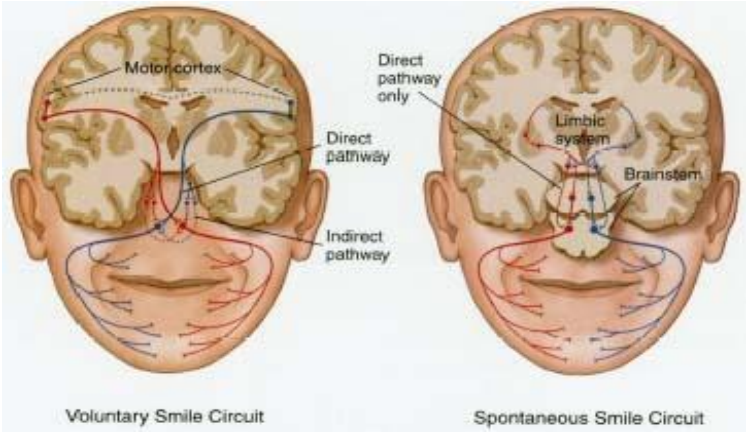


Figure 1-1: The two types of smile circuits [18].

There are some early theories about how emotions are generated. In the 1840s, James Lange was one of the first who studied the relationship between emotion and body physical changes; he proposed that emotions are dependent on two factors: physical changes and the understanding of these changes. He assumed that physical changes happened first, followed by their interpretation, which when combined lead to the creation of emotion. According to his theory,

the body first sends out chemical messengers (like adrenaline), which will cause physical changes, then the brain senses these physical changes and produces emotions [7].

The Cannon-Bard theory, developed by Walter Cannon and Phillip Bard in the 1920s [8], stated that emotions are produced by the nervous system without the need for physical feedback from chemical messengers. They assume that the body senses an event and sends a message to the brain. The brain responds by simultaneously sending messages to the cortex and the hypothalamus. The cortex is responsible for creating emotions, whereas the hypothalamus controls automatic body responses by producing chemical messengers. These messengers are responsible for the behaviour and physiological responses such as crying, fast/slow breathing and shaking [8]. In the 1960s, Stanley Schacter and Jerome Singer combined the James-Lange and Cannon-Bard approaches and produced 'The Schacter-Singer model' [9]. According to their model, both brain processing and physical changes are needed to fully experience any emotion. In their model, information from the environment, body feelings and previous experience all contribute to produce emotion [6].

1.2 Emotion Expression Models

Emotions play an important role in everyday life, such as when communicating, behaviour and creativity. This can impact learning and employment as well as personal relationships. Basically, human's express emotions using the following three models. They are,

I. human speech,
II. body poses,

III. and facial expressions.

Speech is not confined to the literal meaning of words. Research has found that humans change their tone of voice based on their emotional state [10]. The ability to recognise these signals will differ based on language, country and other factors. Body pose involves non-verbal (visual) communication. This model has been studied by social scientists who try to connect body pose to specific emotions. Some of the studies suggest that walking with the head held high is associated with positive emotions, whereas walking with the face down suggests that a negative emotion is being experienced. Figure 1-2 highlights the relationship between body pose and emotion, as suggested by Konrad Schindler [11]; multiple body poses are connected to display single emotions. The third model for expressing emotion is facial expressions this is another form of non-verbal (visual) communication. They are emotional states represented by the movement of facial muscles. This is discussed further in Section 1.5.

Figure 1-2: Emotions represented by body poses [6].

1.3 Emotion Classification

How many emotions exist? This question has not been fully answered. This is due to the complexity of emotions which are influenced by multiple factors. For example, people from different regions experience the same emotion in different ways. Furthermore, people from the same region will experience emotions in different degrees, qualities and intensities which will add another dimension of complexity to the emotion classification process. Moreover, multiple attempts have been made to categorise emotions. The main theories in this field are:

I. Ekman's List of Basic Emotions (1972),
II. Plutchik's Wheel of Emotions (1980),
III. Parrott's Classification of Emotions (2001),
IV. Other theories.

1.3.1 Ekman's List of Basic Emotions (1972)

Ekman completed research on different cultures, finding six common emotions related to facial expressions [3]:

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

In order to identify facial expressions, Ekman's methodology was divided into two main parts. Firstly, he described a situation and asked individuals to choose the facial expression that best fits it, which identified different people's reactions to the same situation. Secondly, he asked the individuals to identify different facial expressions by showing them a picture. Moreover, he asked them to simulate facial expressions based on this picture in order to classify emotions and trigger different facial muscles used to express facial emotions [3]. In the 1990s, Ekman added extra emotions to the six basic emotions described above but stated that not all of these can be encoded via facial expressions:

- Amusement
- Contentment
- Embarrassment
- Excitement
- Guilt
- Pride in achievement
- Relief
- Satisfaction
- Sensory pleasure
- Shame

1.3.2 Plutchik's Wheel of Emotions (1980)

The 'wheel of emotions' was introduced by Robert Plutchik following Ekman's work. This model shows how different emotions can be mixed together to create new emotions. According to Plutchik, basic human emotions consist of four pairs of opposites [12]:

- Acceptance and disgust
- Fear and anger
- Surprise and anticipation
- Sadness and joy

As shown in Figure 1-3, Plutchik adopted the new approach of arranging these eight primary emotions; he suggested that all human emotions are derived from these. Plutchik's 'wheel of emotion' is a 3D model where the adjacent and opposite emotions are connected. For example, optimism is shared by both anticipation and joy, whereas fear and surprise are both linked with awe. According to Plutchik, adjacent emotions blend together to form more complex feelings listed on the outer rim of the emotion wheel. From looking at the wheel of emotion, it seems that from a logical point view, love involves at least some elements of joy and acceptance. On the other hand, contempt will involve components of both anger and disgust [12].

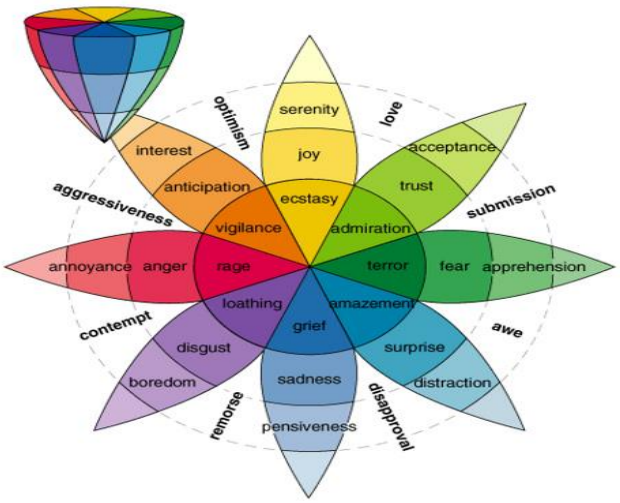


Figure 1-3: Plutchik's wheel of emotions [10].

1.3.3 Parrott's Classification of Emotions

In 2001, a theory was presented by W. Gerrod Parrott consisting of a tree structure classification for deeper emotions; the tree includes over 100 emotions, as shown in Figure 1-4. The tree comprises of six primary emotions, 25 secondary emotions and more than a hundred tertiary emotions [13].

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
	Sympathy	Pity, sympathy
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Figure 1-4: Parrott's classification of emotions (2001) [24].

1.3.4 Other Theories

There are many theories which attempt to classify emotion. Figure 1-5 represents 'A Selection of Lists of Basic Emotions' presented by Ortony and Turner in 1990, showing different theories and the corresponding basic emotion related to it [14].

Theorist	Basic Emotions
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Arnold	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Ekman, Friesen, and Ellsworth	Anger, disgust, fear, joy, sadness, surprise
Frijda	Desire, happiness, interest, surprise, wonder, sorrow
Gray	Rage and terror, anxiety, joy
Izard	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
James	Fear, grief, love, rage
McDougall	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
Mowrer	Pain, pleasure
Oatley and Johnson-Laird	Anger, disgust, anxiety, happiness, sadness
Panksepp	Expectancy, fear, rage, panic
Tomkins	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Watson	Fear, love, rage
Weiner and Graham	Happiness, sadness

Figure 1-5: Theorist and 'basic emotions' [14].

1.4 Computer Vision and HCI

Humans use their vision and sense of touch to analyse the surrounding environment. Computer vision is a science that aims to simulate these capabilities when analysing objects from the surrounding environment by using images and videos. Furthermore, computer vision can be defined as "*a field that includes methods for acquiring, processing, analysing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information*" [15]. Computer vision's main concern is developing theories for building artificial intelligence systems that retrieve information from image data which can be articulated into multiple modalities such as videos, multiple camera views, medical image devices, a depth camera and more.

Computer vision hierarchy consists of three levels [16]: low, middle and high-level vision. The low-level vision includes feature extraction from images like

edges, corners or computing optical flow for motion analysis. Middle-level vision uses features obtained from low-level vision to do object recognition, motion analysis and 3D reconstruction. Finally, high-level vision is based on the interpretation of the information obtained from middle-level vision. Furthermore, high-level vision directs how tasks should be performed for both middle and low-level vision. The interpretation may include the conceptual description of a shape, motion vectors, activity or behaviour [16]. Computer vision applications are most commonly used in the following fields [15]:

1) Robotics
2) Medicine
3) Security
4) Transportation
5) Industrial automation
6) Image/video databases
7) Human computer interface

Robotics uses computer vision for analysing the surrounding environment and to recognise objects, things and humans. The medical field uses computer vision to detect and quantify diseases from images obtained through magnetic resonance imaging and ultrasounds as well as other medical imaging devices. Security uses computer vision to detect motions and analyse faces for face recognition applications. Transportation uses it in autonomous vehicles to identify objects, roads, vehicles and humans. Industrial automation uses computer vision to speed up the process whilst maintaining high-quality products without the need for human interference. Finally, image/video databases use computer vision to sort and retrieve media from large databases [15].

Human computer interface is a combination of HCI models and computer vision. Computer vision deals with understanding images obtained from different resources whereas HCI models try to enhance a system interface using more human-like communication techniques. HCI uses computer vision techniques as enabling techniques to achieve more human-centric interfaces/systems. Human computer interface can be classified into four main categories [17]:

I. Large-scale body movements
II. Gesture recognition
III. Gaze detection
IV. Affective human computer interaction (emotion recognition)

Large-scale body movements can be measured through head, arms, torso and legs and can be used in a variety of applications in human computer interaction (HCI). In [18] there are a set of applications based on body movement such as smart surveillance systems, choreography of dance and more. Furthermore, body movements can be used to analyse movements to learn, operate machines or commands. This can be achieved by applying a motion algorithm which includes gesture recognition and motion analysis [18].

Gesture recognition is a mathematical representation of objects, human gestures and environments retrieved from images, videos or both. An example of gesture recognition is a hand gesture. Examples presented in [19] use hand gestures to write on surfaces and then transfer the writing to a display and in [20] they propose a framework for contorting robotic arms using hand gestures. Finally, gesture recognition can be used in identifying body movement recognition, gaze recognition and emotion recognition [19].

Gaze detection is research done to identify the direction of eyes in space. Research carried out on gaze detection can be classified into two main categories: wearable or non-wearable, and infrared or appearance model [17]. Gaze detection systems use a mix of these categories. Wearable or non-wearable represents the state of devices used in gaze detection, where researchers found that wearable infrared systems have more accuracy than non-wearable [17]. Infrared systems use the ‘red-eye effect’ method which detects eye location by reflecting infrared rays into the face, which converts the eye colours into red colour; this simplifies the detection of the eye location. After detecting eye location, a set of measurements and algorithms are applied to detect the eye corner point, shape and motion. In [21] they used an infrared camera to detect the eyes and then used approximate measurements to detect and analyse the location of the face, as shown in Figure 1-6.

Figure 1-6: Eye detection using infrared [29].

The appearance model uses gesture algorithms to detect the eye pupil. Machine-learning techniques can be used to train the system, such as principal component analysis (PCA) [22], which was used with a large database that contained images of the eye and non-eye images. A large dataset was used to train the system to distinguish between the eye and non-eye objects. The author

of [23] shows an example of gaze detection by first detecting the face using the Paul Viola and Michael Jones algorithm [24] and then applies gradients algorithms to detect the eyes [23].

Affective human computer interaction (AHCI), or emotion recognition, is the machine's ability to adopt human/user behaviour and use this information to improve or develop interfaces [17]. Moreover, AHCI tries to combine emotion recognition with technology to create more natural interaction and communication. AHCI applies enabling techniques to technologies to produce the impression of experiencing emotions [17].

Emotion recognition can be categorised into three main areas [17]:

I. Body language
II. Audio
III. Facial expression

Recent work done on body poses that are related to emotion recognition attempts to categorise specific poses with their corresponding emotions. As shown in Figure 1-2, Konrad Schindler [11] analysed an image of a body pose using the Gabor function to find the body orientation and then used principle component analysis (PCA) to match the body pose to a specific template that is used later in the classification process [17].

Audio is used in emotion recognition to investigate how to retrieve emotional information from vocal modalities. As previously mentioned, audio-based HCI analyses human speech through multiple algorithms. In [25] using different algorithms to analyse speech frequency, such as Mel Frequency Cepstral Coefficients (MFCCs), and apply them on an Emotional Dyadic Motion

Capture database (IEMOCAP). Gaussian Mixture Models (GMMs) are used to classify and cluster voice to specific emotions based on the features extracted from MFCCs [17].

1.5 Facial Expressions

"*The face, they say, isn't the mirror to emotions it's been held out to be*" [26]. Facial expressions are a part of non-verbal (visual) communication and considered to be emotional states represented by facial muscle movement. Essentially, when a person experiences an emotion, different inner organs trigger facial muscles to change to express emotions. Moreover, facial expression provides information about humans, the way they react to specific situations and sometimes tells us something about the situation itself.

Humans can embrace facial expressions voluntarily or involuntarily. Voluntary control often occurs when individuals are in social situations where they may feel obligated to simulate emotions; voluntary facial expressions are generated from the brain. Involuntary or 'spontaneous' facial expressions are caused by situations which trigger emotions without awareness. Involuntary movements are produced by the brain via the cortical route which is derived from the sub-cortical route. This explains why some people do not realise that they are expressing facial expressions; their body and brain sense the circumstances of the environment and produce facial expressions based on it [27].

Based on Ekman's research [3] in the field of automatic facial expression analysis there are six basic emotions (fear, sadness, disgust, anger, surprise and happiness) generated using facial expressions. However, as mentioned before, these basic expressions represent only a small set of human facial expressions.

1.5.1 Facial Expression Measurement

A facial gesture can be characterised by retrieving information from the face and facial expressions. The Facial Action Coding System (FACS), developed by Ekman and Friesen [28], is one of the most commonly used systems for facial behaviour analysis in psychological research. Another approach is to use geometrical terms to describe facial motion which is known as facial animation parameters (FAPs) [29].

1.5.1.1 Facial Action Coding System (FACS)

FACS is a system used to categorise human facial movements based on their appearance on the face. FACS was originally developed by a Swedish anatomist named Carl-Herman Hjortsjo [30]. Thereafter it was adopted by Ekman and Friesen [28] in 1978. According to Ekman's work, FACS is a representation of facial expressions constructed by muscles or a group of muscle movements. These movements are called 'Action Units' (AUs). By observing AUs, FACS can detect and measure facial expressions. Ekman has identified 46 AUs, where a single AU or group of AUs can be used to represent facial muscles.

Initially, FACS was founded to train humans (coders) to encode facial expressions using AUs. Figure 1-7 displays some AUs and action unit descriptors (AUDs). AUDs can be defined as movements that may involve the actions of several muscle groups used to activate the AU (e.g. a forward-thrusting movement of the jaw).

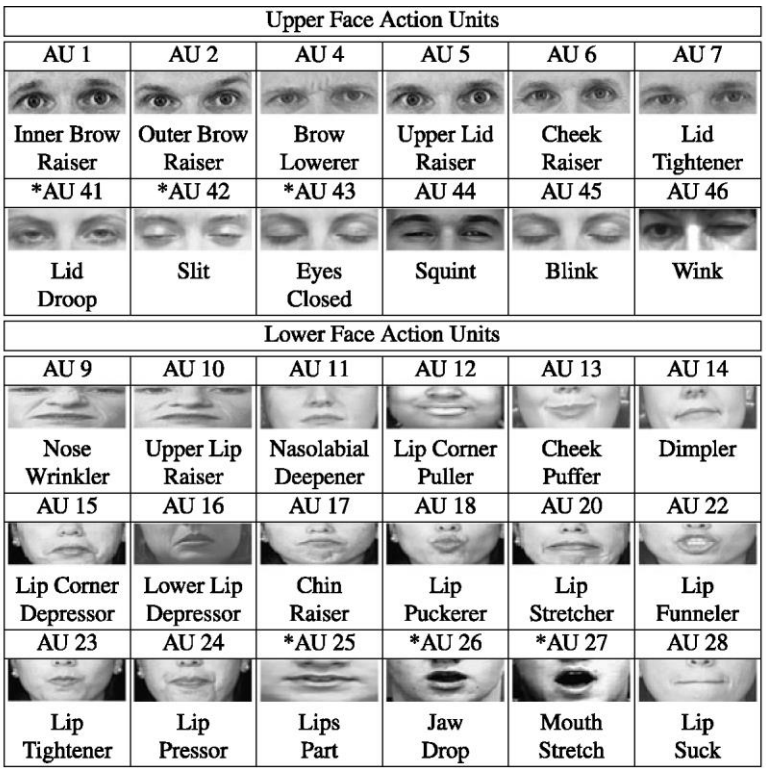


Figure 1-7: FACS, Action Units, Action Descriptors [31].

Ekman's FACS manual is over 500 pages in length and provides the AUs, as well as their meaning FACS. Using FACS, humans (coders) can construct and deconstruct facial expressions into specific action units [32], where action units are independent of one another. Furthermore, AUs can be used in decision-making processes to recognise basic emotions, where each emotion can be divided into a group of AUs.

Describing facial expressions is challenging. According to Ekman, 100 hours were needed to train coders to use FACS accurately and reliably. Furthermore, coding time can be expensive. As a result, only a small number of coders can successfully code facial expressions using FACS [33].

In 1981, Ekman presented the EMFACS (Emotional Facial Action Coding System) [31]. According to Ekman, EMFACS is more economical in comparison

to FACS as coders are not required to detect each muscle change; instead, they decide if a group of changes is associated with specific emotions. EMFACS is a shorter version of FACS that takes less time to analyse and produces the same results, as shown in Table 1-1 [34].

Table 1-1: EMFACS and AUs [34].

Emotions	Action Units
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26
Fear	1+2+4+5+7+20+26
Anger	4+5+7+23
Disgust	9+15+16
Contempt	R12A+R14A

1.5.1.2 Facial Animation Parameters (FAPs)

One of the applications of the FACS model was produced by group MPEG-4 (Moving Picture Experts Group) called Facial Animation Parameter (FAP), as a part of Face and Body Animation (FBA). FAP is responsible for characterising animation to look and behave like humans. The development of FAP is to allow a definition of both facial shape and texture. Furthermore, FAP and FBA play a role in producing facial expressions, emotions and speech pronunciation in the animation environment. The main purpose of FAP is to animate human or human-like characters in 3D environments [29].

FAPs represent a complete facial action set founded on the study of minimal facial actions related to muscle actions. Compressing or lowering

eyebrows and opening or closing the mouth are examples of these facial actions [29]. In order to make the facial model more consistent, facial animation parameter units (FAPUs) were created [35]. FAPUs correspond to distances between some key facial features as shown in Figure 1-8 (a). FAP consists of feature points spread over the face leading to the construction of a mesh. This is known as facial definition parameters (FDPs). These feature points are used to produce animated visemes and facial expressions as shown in Figure 1-8 (b) [29].

According to [36], FAPs signify 66 displacements and rotations of the feature points from the neutral face position. Accordingly, FAPs are used in computer generated imaginary (CGI) which investigates ways to design virtual faces and human facial expressions in the animation environment [37].

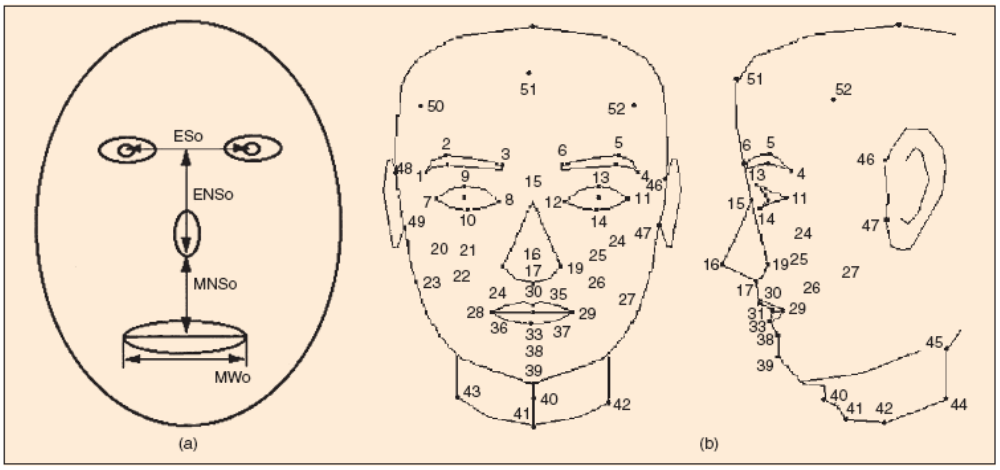


Figure 1-8: The Facial Animation Parameter Units (FAPUs) and the Facial Definition Parameter (FDP) set defined in the MPEG-4 [29].

The main difference between FAP and FACS is that FAP defines facial muscle movement by creating a displacement of a pre-defined location on the face, taking into consideration the natural face [38]. On the other hand, FACS

describes facial expressions using AU representing muscle movements which cannot decompose into basic ones [39].

1.6 Objectives

The objectives of this research are to work towards the development and understanding of the facial expressions and identify different aspects of it. More specifically, our objectives are to:

- examine how much facial expression is encoded within the facial dynamics.
- classify gender based on the dynamic characteristics of the smile without taking into consideration any texture data.
- identify the relationship between facial features with real and posed smiles. Furthermore, how these facial features contribute to formulating each type of smile.
- identify the possibility of using facial expression as biometric using only the dynamic characteristics of the facial emotions.

1.7 The Importance of this Research

In this research, we cover different aspects of emotion detection and analysis. In chapter 4, we present an automated algorithm to detect emotions using the FACS system. This is done by detecting face and facial muscles movement to identify 19 AUs which correspond to one of the six basic emotions identified by Ekman (happiness, surprise, sadness, fear, anger and disgust). The importance of this research is the ability of machines to understand emotions and how the user feels. Furthermore, it can be used as an assessment tool for evaluating FACS coder or as analysis tools for interviews.

In addition, we study real and fake smiles. Chapter 5 shows a non-invasive way to analyse real and fake smiles and identify related distinguishing facial features. This research has been carried out and proved by many psychological studies and our results show that it is approved with the finding of the psychological studies. Our results indicate that eye movement is the distinguishing feature between real and fake smiles. Moreover, we identify different statistics that show the weight of each facial feature in real and fake smiles. The importance of this research is to identify fake and real smiles and the corresponding weight for each of the facial features and it gives the machine the ability to understand another dimension of emotions.

In addition, we show that emotions not only represent human feelings, but they can determine their gender and identity. In chapter 6, we study the smile expression for clues for gender variances. We chose the smile expression since it is considered one of the most universal facial expressions among different cultures. Our approach introduces a new algorithm to analyse the dynamic facial feature (texture-less feature) and use it to classify gender. As result, we can classify 86% based on smile dynamic only. The importance of this research is to show a fast and robust way to identify gender using a minimal set of features.

In chapter 7, we study the smile dynamic in clues for a biometric. We present an automated algorithm to identify emotions' signature using smile expression. Our results indicate two key points: the uniqueness and the similarity of smile expression among humans. This research focuses on the dynamic features of smile and does not take into consideration any texture data. The importance of this research is to introduce a new biometric that can be used to

identify humans using the facial expression or can be used as a further stage for other biometrics.

1.8 Contributions

The outcome of this research and the main contributions presented in this thesis are summarised as follows:

1) In chapter 3, we enhance the detection of facial features by applying regions of interest (ROI). The result shows high accuracy when compared to the original algorithm with the ROI applied. Furthermore, it reduces the percentage of false positive for detecting the mouth, eyes and the nose. This research has been successfully published in 16th UK Workshop on Computational Intelligence [40].

2) In chapter 4, we present a fully automated system to detect facial expressions and action units. This framework is based on facial muscle movement only. This system shows a high detection rate using the minimal set of features.

3) In chapter 5, we study the relationship between facial features in terms of real and fake smiles. In this research, we undertake a statistical analysis of the smile to determine which facial features weight are prominent in identifying genuine and posed similes. As a result, we find that the eye contains the most information in distinguishing between a fake and a real smile. Furthermore, we show the weight of each facial feature in genuine and poses smiles.

This research has been accepted as a conference paper in cyberworlds conference 2018 and has been requested to submit an extended version of this research to the conference special issue journal.

4) In chapter 6, we study the dynamic characteristics of the smile which may provide clues for gender. In this research, we present a novel algorithm that uses a set of spatial, geometric and flow parameters which represent only the dynamic features of the smile. Using these features, we can classify gender with a high classification rate using a machine-learning algorithm. This research has been successfully published in the Visual Computer journal [41].

5) In chapter 7, we study the possibility of using the dynamic characteristics of the smile as a biometric. In this research, we present a novel algorithm to identify the unique characteristics of the smile. The results show the degree of similarity among subjects using the dynamic features in the smile. Additionally, the result shows that the human smile has unique characteristics that can be used as a biometric.

1.9 Datasets

To test our proposed methodologies, we mainly apply our algorithms on two main datasets; these datasets are used to evaluate the accuracy and the correction of our approach. These datasets are the CK+ [42] and the Multimedia Understanding Group (MUG) datasets:

I. The Cohn-Kanade AU-Coded Facial Expression Database contains two versions: CK and CK+. CK includes 486 sequences from 97 posers. Each sequence begins with natural expression and proceeds to a peak expression. The peak expression for each sequence is fully FACS coded and emotion labelled. CK+ includes both posed and non-posed (spontaneous) expressions. For posed expressions, the number of sequences is increased from the initial release by 22% and the number of subjects by 27%. Furthermore, these are also fully FACS and emotions coded.

II. The MUG dataset [43], consists of image sequences of 86 subjects performing a non-posed facial expression (laboratory induction expression). The environment was controlled where subjects were sitting in a chair in front of one camera, a blue screen background and fixed light sources. The camera was able to capture images at a rate of 19 frames per second. Each image was saved with a jpg format, 896×896 pixels.

The MUG database consists of 86 participants (35 women, 51 men), all of Caucasian origin, between 20 and 35 years of age; men are with or without beards. The subjects are not wearing glasses except for seven subjects in the second part of the database. The images of 52 subjects are available to authorised internet users. Twenty-five subjects are available upon request and the remaining nine subjects are available only in the MUG laboratory. The MUG dataset is not FACS coded.

1.10 Outline of the Thesis

The remainder of the thesis is organised into six chapters. Chapter 2 is a literature review for each following chapter. It shows different techniques used in each related chapter. It also indicates gaps in knowledge that different research has tried to tackle. Chapter 3 presents a framework to enhance the facial features detection. Chapter 4 presents a framework to analyse action units and emotions using facial muscle movement. Chapter 5 presents research for distinguishing real and fake smiles and presents a statistical analysis of the facial features related to each smile. Chapter 6 presents a new algorithm to identify gender using smile dynamics. Chapter 7 shows a new biometric algorithm to identify subjects using the dynamic features of the smile. Lastly, in Chapter 8 we conclude this thesis and discuss possible limitations and future work.

2 Literature Review

Emotions play a very important role in non-verbal communication as they provide an insight into human feelings and interactions [44]. Moreover, as humans have become heavily dependent on technology, using devices such as computers, phones and tablets, it would be useful if computers could synthesise human facial expressions [44], understand them and identify different aspects of emotions that contain additional characteristics besides facial expressions. According to [45] facial expressions do not reflect only emotions, but describe other signs such as mental activities, social interaction and physiological signals.

There is a massive range of applications of facial expressions analysis. These include image understanding, psychological studies, medical[46, 47], behavioural science[34], face image compression and face animation [48]. Additionally, facial expression analysis can enhance the interaction between humans and technology by tracking, analysing and understanding human emotions. Furthermore, it aids in developing high-quality software [15] and advances computers' power to understand humans emotions more deeply.

In this research, we study different aspects of facial expression detection and analysis from different field viewpoints. Such fields include computing, physiological, psychological and social where each field contribute to developing and motivating each research carried in this thesis.

This chapter structured as follows: Section 2.1 represents the literature review for Chapter 4 which covers the Facial Action Coding System (FACS) detection and analysis using facial movement. Section 2.2 represents the literature review for Chapter 5 which covers fake and real smiles and their relation

to facial features. Section 2.3 shows Chapter 6's literature review in identifying gender using smile expression. Section 2.4 shows the literature review for Chapter 7 which covers the use of facial expression as a biometric.

2.1 FACS Detection and Analysis

Facial expression universality has already been recognised for six basic emotions by Paul Ekman [3]. However, the number of emotional facial expressions used in daily life is much higher, either by combining basic emotions together to produce new emotions or other categories of emotions as described previously in Section 1.3. The majority of research carried out on emotion recognition using facial expressions tries to identify the six basic emotions, a subset of those and the action units related to the six basic emotions.

As described previously in Section 1.5.1, the FACS is one of the most well-known systems used to analyse facial expressions. Using the FACS system to identify facial expressions by using action units, which represent muscle or groups of muscle movement, a total of 46 AUs were presented by the FACS system.

Automatic action unit detection has received a lot of attention over the years, specifically, in terms of analysing facial expressions and generally, classifying action units done by extracting features from the face which is done using one or more of three models: texture, geometric distance and motion data.

The texture model can also be referred to as template matching or using an exemplar of the object. The reason for using appearance models is that objects look different under changes in lighting, colour, direction and at different scales. Furthermore, it describes the texture of the face caused by the expression

or appearance of new features on the face (such as wrinkles and furrows) [49]. The geometric model measures the movement of points through specific areas. Moreover, the geometric model tracks landmarks that describe the shape of the face and its components, such as the mouth or eyebrows. Some research uses a hybrid model (appearance and geometric) in order to get higher accuracy [50] in the detection phase.

There are a lot of challenges facing AU detection which include environment factors such as: illumination, pose and occlusion, and individual factors such as: shape, texture, scale and behaviour. To address some of these challenges, various approaches have been proposed for detecting facial expression and action units. In this section we describe research done in this field.

A lot of state of the art research carried on identifying the facial expressions and action unit detection using either Neural Network (NN), Convolutional Neural Network (CNN) or Deep Convolutional Neural Network (DCN) which reported very high classification rates. In [51] they use a set of techniques to extract features for detecting facial expressions; such techniques include image resize, convert to greyscale format, face detection using viola-jones algorithm and landmarks detection using D-lib ML toolkit. Additionally, they use Central Binary Patterns (CBP) for computing additional facial features and use PCA to reduce the dimensions of the features detected. For classification, they design a Convolutional Neural Network (CNN) with 10 layers to classify the facial expressions where they use CK+ dataset for training the CNN. As a result, they gain 95% correct classification for the six universal facial expressions.

In [52] they proposed a deep conventional network to identify facial expressions and relation traits, i.e. dominant, competitive, trusting, warm, friendly, involved, demonstrative and assured. This done by training a fully connected DCN on two new datasets namely, Exp-W (a large-scale face expression dataset in the wild) and a new dataset for relation traits. Exp-W contained a total of 91000 images collected form the web and is considered to be the wild dataset, the new dataset contains a total of 8016 images collected from the web and movies which is manually labelled with pairwise relations. These datasets are used to train a DCN to identify the facial expressions and relation traits. As a result, they reported 98% correct classification for facial expressions and 71% on relation traits.

In [53] they detect facial expression using CNN, where they proposed a hybrid form from two convolutional neural networks namely, PHRNN (part-based hierarchical bidirectional recurrent neural network) and MSCNN (multi-signal convolutional neural network). The PHRNN is responsible for analysing temporal features which contain facial variations and facial geometries. The MSCNN is responsible for analysing spatial features which include analysing appearance information from still images. Using a hybrid model form the two networks they reported 98.5% classification on the CK+ dataset.

In [54] they try to identify facial expressions using CNN. This was done by using a set of pre-processing techniques before initialising the training of the CNN such techniques include face rotation, cropping, downsizing and Intensity normalisation. Additionally, for training the CNN they use a set of public datasets which include CK+, JAFFE and BU-3DFE. As result, they reported a 96.7% on the CK+ dataset.

In [55] they present FaceNet2ExpNet which is a fully connected convolutional layer used to train an expression recognition network from static images. This was done by applying two-stage training algorithms (pre-training stage, refining stage). Pre-training stage done by training the convolutional layers on the Expression-Net and Face-Net model. Face-Net model extract feature which contains expressive information about human faces which will be used to identify facial expressions using in Expression-Net model. Refining stage consists of appending the fully connected layers to the trained convolutional layers (Expression-Net and Face-Net) where they train the whole network jointly using the cross-entropy loss. For evaluating their proposed network, they use public datasets such as CK+, Oulu-CASIA, TFD, and SFEW. As a result, they reported 98.6% correct classification.

In [56] they present a framework for detecting facial expressions and identify pain and non-pain faces. This was done by combining CNN from the VGG-face model with Long Short-Term Memory (LSTM) model. Their method consists of two phases: pre-process and training phase. Pre-process phase contains landmarks detection and image alignment. Training phase which use the output of the pre-processing phase to train the VGG-face and LSTM to extract features to identify the facial expressions and pain intensity in the face. For testing facial expressions detection, they use CK+ dataset and reported 97.2%. For identifying pain and non-pain they use UNBC-McMaster Shoulder Pain Expression Archive Database and reported 93.3 % with leave-one-subject-out validation model.

In [57] the authors attempted to identify four facial expressions (happiness, sadness, anger and null) by tracking facial landmarks. They started with face normalisation (angle and size), by locating the face manually in the image and resizing it to 256*256 before converting it into a greyscale image. Then, they manually pre-defined five feature regions that included eyebrows (left and right), eyes (left and right) and mouth which are used as input data for the tracking algorithm.

These features are determined by manually selecting their location in the first frame and using a tracking algorithm to track the movement of these points through the image sequences. The Lucas and Kanade optical flow algorithm [58] is used to track the points, which computes the movement vector for each facial point between adjacent frames in the video. The output is 26 feature vectors with eight vectors for the eyebrows, eight for the eyes and 10 for the mouth. Figure 9 represents the process used to identify the facial landmarks and track them using optical flow [57]. Figure 2-1(a) identifies the landmark; Figure (b) identifies and crops the five feature regions, Figure 2-1(c and d) shows the tracking of the landmark movement using the optical flow algorithm.

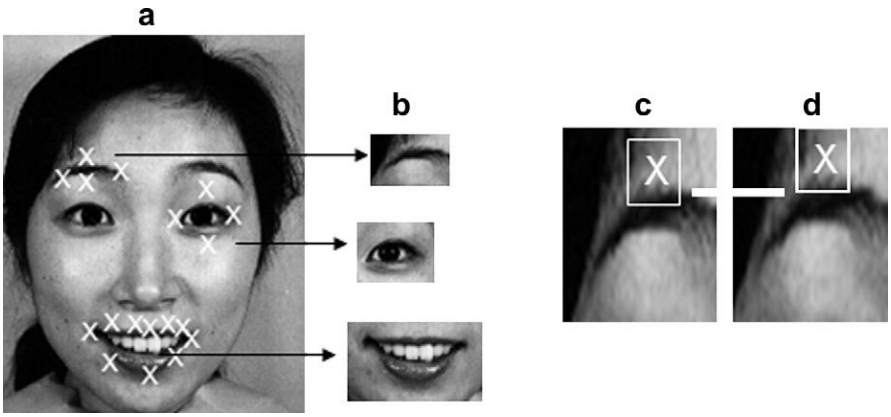


Figure 2-1: Feature detection: (a) feature points, (b) cropped regions, (c, d) the displacement of a feature point [49].

The classification process tries to classify four basic emotions as mentioned previously: happiness, sadness, anger and null. The classification process compares angle and magnate of the feature vector produced by the input for the pre-calculated feature vector on the Kanade, Cohn and Tian database [65]. The results found were that 83.33% of the image sequences were classified in either of the four emotion categories, while 16.77% were not [57].

In [59] they proposed a hybrid system for detecting action units based on appearance features and motion data. They analysed facial feature appearance by detecting facial landmark, then extracting motion data using rigid and non-rigid registration with contracting quad-tree decomposition which is responsible for defining interesting face regions related to that AU. For classification, they used a gentle boost and the hidden Markov model (HMM). Their system can classify 30 facial action units with an average of 94.3%.

In [60] they analysed texture feature to detect facial actions in spontaneous expressions. This was done by applying Gabor wavelets. Since Gabor wavelets produce a large number of features they used AdaBoost for features selection. Finally, they used a support vector machine to classify these features into 21 AUs. They applied their proposed method on the Cohn-Kanade and Ekman-Hager database and obtained a mean of 91% agreement with human FACS labels provided.

In [45] they proposed a system to identify facial expression in crisis environments. This is done by applying an Active Appearance Model (AAM) which extracts shape and texture features. Then, using Facial Characteristic

Points (FCP), which are used to compute a set of suitable parameters that give a good description of the appearance of the facial features extracted by the AAM. For classification, they used a two-fold support vector machine where the results have an average of 90.4% for detecting 24 AUs.

In [61] they presented results applying AAM-derived facial representations, to identify facial action recognition. This is done by applying a set of methods: first, 2D shapes of a 2D AAM. Second, a 2D appearance which is a representation of the appearance of the face. Finally, a 3D shape where the position in 3D of the face and its facial features has been estimated. For classification, they tested two main classification algorithms: nearest neighbour (NN) and support vector machine (SVM). The result was to test the identification on four AUs (AU1, AU1+2, AU4, AU5) and they found that both shape and texture representation is significantly beneficial for recognition performance. Furthermore, shape features have a large role to play in facial action unit recognition.

In [62] they proposed a multistate feature base for detecting AU. This was done by using a set of geometric distances and a set of methods to identify wrinkles caused by AU formation. For classification, they used a neural network (NN) to identify six AUs and a combination of these AUs as a result, they correctly classified 95%.

In [63] they proposed an automated system for detection of AU. This was done by applying a set of geometric distances and applying Gabor wavelets in a predefined region of interest (ROI). For reducing the number of features extracted using Gabor wavelets they used a gentle boost. For classification, they used a hybrid system combining an SVM and an HMM. This approach has been applied

previously on speech recognition. As a result, they correctly classified 23 AUs with a mean of 66%.

In [64] they designed a system to classify six basic facial expressions (happiness, surprise, sadness, anger, fear and disgust) based on motion data. This was done by analysis motion by applying an optical flow algorithm on eyebrows, eyes, nose and mouth. For facial expressions classification, they used a look-up table.

In [65] they presented an automated system for detecting action units. In their work, they applied a set of methods to detect six action units that include holistic spatial analysis they got 89% correct classification. With Principle Component Analysis (PCA) on a greyscale image to detect wrinkles, they got 57%; with template matching and optical flow they got 85% correct classification. When combining the three methods they got 92%.

In [21] automatic identification of the face region was done using IR illumination. As shown in chapter 1, Section 1.4, Figure1-6, the purpose is to identify the eye location first then use it as a basis to determine other face regions and face segmentation. Face segmentation includes defining the eye location and drawing a line between the eyes. Furthermore, another line is drawn parallel to the first one in a way that divides the area between the first line and the edge of the face image, as shown in Figure 2-2. The segmentation divides the face into six main parts. The motion vector for each part is calculated because of the assumption that most motion vectors will be in these parts and each part will have its own and unique direction. Lucas and Kanade's optical flow algorithm [58] is used to compute motion vectors and produce "source vector sets", which are a

collection of vectors, representing motion and deformations in the face caused by the emotion representation [21].

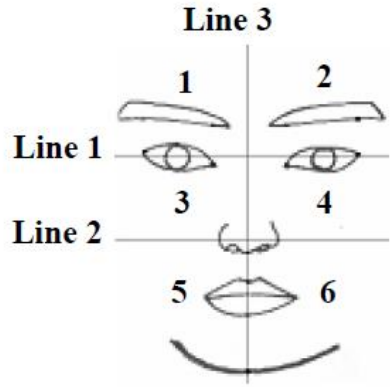


Figure 2-2: Face segmentation into six parts [21].

In order to compute the source vector, the authors [21] define it as: "*A series of facial images (is) chosen which its motion vectors show facial expression correctly similar to Bassili description. Then incorrect vectors are eliminated and for others a pair of vector angle and its area is saved*" [21]. Bassili's description of emotions [66], shown in Figure 2-3, shows which facial expression cues are related to specific emotions and how to represent the emotions using motion vectors. The main advantage of the source vector is the impact of the correct vector, considered to be a higher property and the elimination of the noise vector. The classification phase is done by estimating similarities between the source vectors and the extracted motion vectors. This work found 83.3% correct classification on the Cohn-Kanade facial expression database [67].

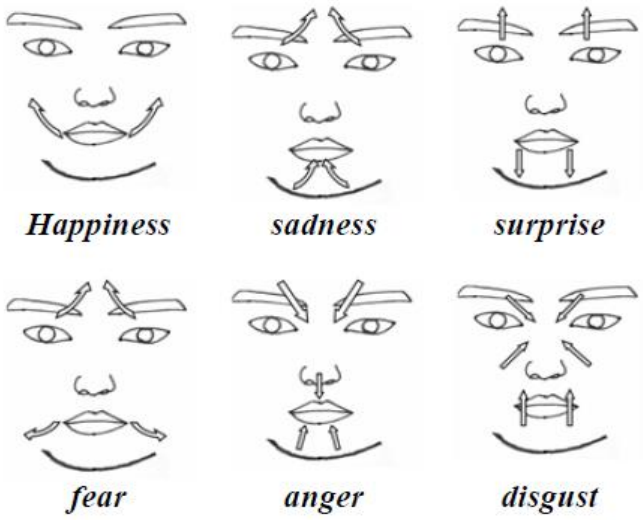


Figure 2-3: Facial expression identified by facial features movement as suggested by Bassili [58].

In [68] they attempted fully automated facial action coding for spontaneous expressions. Their system automatically detected frontal faces using a modified version of the Viola-Jones algorithm [24] and detected 20 action units for each frame. Furthermore, face images were rescaled to 96*96 pixels and an automated eye detection was applied in order to align images where the distance between the eyes was roughly 48 pixels.

After detecting the face and the eyes, a set of Gabor filters with eight orientations and nine spatial frequencies were applied in order to extract face texture as shown in Figure 2-4. Since the number of features extracted from the Gabor function was very high, this gave 9x8x48x48 = 165,888 possible features. A subset of these features was therefore chosen using AdaBoost. This method was used on the CK+[42] database and SVM in order to classify the AUs activated in the video or real-time environment. The classification of AUs from the AdaBoost classifier, based on spontaneous emotions including head movement and pose, was 93% correct [68].

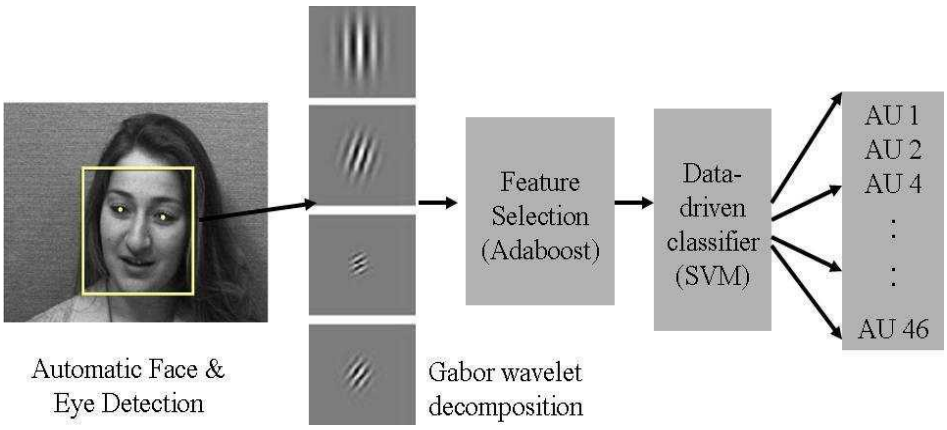


Figure 2-4: Proposed framework by [60].

In [69] the authors recognised facial expressions using geometric information (landmarks) and analysed them using Gabor functions. In their research, they used a database containing 213 images of females. The images were rescaled and cropped so that the eyes were located roughly at position 60 pixels from the right side of the image. The images were then resized to 256*256 pixels. The method was divided into two. The first used 34 facial points (selected manually) as shown in Figure 2-5 . In the second method, features were extracted with 2D Gabor transforms, where 18 complex Gabor wavelet coefficients were applied on facial landmarks. In summary, each image was represented with a vector of 612 elements (which is 34 landmarks * 18 Gabor filters) [69].

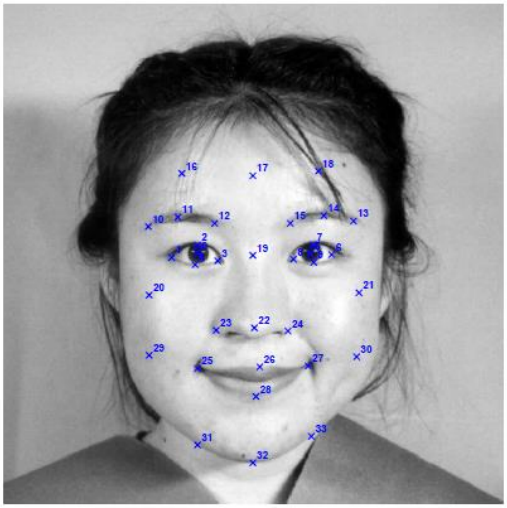


Figure 2-5: 34 facial landmarks represents the facial geometry [61].

The architecture of their system was based on a two-layer perceptron as shown in Figure 2-6. Moreover, there are no interconnections in the first layer between geometric and Gabor wavelet features, because they contain two different types of information. As described before, features (landmarks, Gabor wavelets) are extracted and used as an input to the two-layer perceptron. The first layer performs as a non-linear reduction of the dimensionality of the feature space depending on the number of hidden units. This feature reduction is necessary due to a large number of elements used to represent the vector (612 elements). The second layer makes a statistical decision based on the reduced set of features in the hidden units and consists of seven output units which represent each emotion. The results give 90.1% correct classifications using seven hidden units without the emotion fear. After including fear in their system, they achieved 85.6% correct classifications [69].

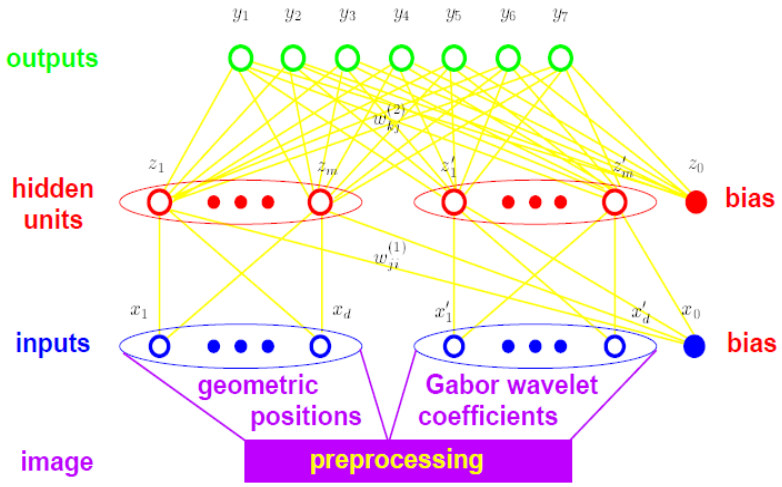


Figure 2-6: The architecture of the proposed system [61].

In [70] they identified facial expressions of people with neuropsychiatric disorders. They started by detecting faces using Viola and Jones [24]. Then they used an Active Shape Model (ASM) [71] in order to find the facial landmarks. ASM works by determining landmarks on the face region manually (in this research they used 159 landmarks) and then this technique is used on each frame to estimate the landmark location for new faces [70].

The ASM outputs a set of landmarks which are used to determine geometric and texture information from images. Geometric features are defined by creating a landmark template using the training data. Starting with the averaged landmark locations as a template, they align landmarks from all training faces to the template and update the template by averaging the aligned landmarks. This procedure is repeated for a few iterations until it aligns with the face (Figure 2-7) [70]. Gabor functions are used with nine different spatial frequencies from 1/2 to 1/32 with eight different orientations to find different kinds

of texture. Both geometric and texture information is found as they convey complementary information.

Figure 2-7: Landmark aligned with face using ASM [62].

Geometric information can be used to detect changes in the face resulting from AU. Furthermore, in that case, texture changes (e.g. appearance of vertical wrinkles between the eyebrows) capture extra information created by the AU. The features from geometry and texture are combined at the classification stage [70].

For the classification phase, AUs are clustered to specific emotions. The appearance of certain AUs is detected by information found from geometric and texture analysis. This data is inputted into Gentle AdaBoost classifiers, which decide if certain AUs occurred or not. Emotion classification is related to the appearance of AUs. Figure 2-8 shows the AU detection rate; the overall detection was 95.9%.

AU No	Description	Rate (%)
AU1	Inner Brow Raiser	95.8
AU2	Outer Brow Raiser	97.8
AU4	Brow Lowerer	91.0
AU5	Upper Lid Raiser	96.9
AU6	Cheek Raiser	93.0
AU7	Lid Tightener	87.0
AU9	Nose Wrinkler	97.5
AU10	Upper Lip Raiser	99.3
AU12	Lip Corner Puller	97.1
AU15	Lip Corner Depressor	99.2
AU17	Chin Raiser	96.5
AU18	Lip Puckerer	98.6
AU20	Lip Stretcher	97.7
AU23	Lip Tightener	96.9
AU25	Lips Part	95.7

Figure 2-8: Action Units and detection rate [62].

The original purpose of this research was to find facial expressions of people with neuropsychiatric disorders and compare them to the facial expressions found in a normal person. They found that some patients with neuropsychiatric disorders use different AU combinations to express their emotions. For example, “normal” people use action units AU6, AU12, AU4 and AU20 to express happiness. Alternatively, patients with neuropsychiatric disorders use action units AU6 and AU12 to express happiness.

2.2 Smile Weight Distributions

Distinguishing between a real and fake smile is still a very difficult problem, which according to [72], is associated with perceptual and attentional mechanisms, where humans pay a lack of attention to a certain cue that would help them in their judgement between these types of smile.

According to[72, 73] there are two types of smiles – Duchenne and non-Duchenne smiles – which are related to enjoyment and non-enjoyment smiles respectively. Recent research shows that spontaneous smiles are associated with enjoyment which is related to the zygomatic major and the orbicularis oculi muscles movement[74, 75], which in terms of the FACS, is known as the Duchenne smile. This smile is identified by the cheek raiser or known as the Duchenne marker (AU6) and movement of the mouth corner upwards (AU12). The Duchenne marker causes the skin around the eyes to crows, closing the eye opening and formulating wrinkles around the eye area. According to a set of research, this movement is related to expressing a genuine smile or is associated with true emotional experience. On the contrary, a non-Duchenne smile, or non-enjoyment smile, is formulated by a movement of the mouth corner upwards (AU12).

Recent research in analysing real (Duchenne) and fake (non-Duchenne) smiles has been carried out from three perspectives: physiological, social experiments and computational studies.

2.2.1 Physiological Studies

Facial expression analysis and detection is also an active research field in psychology where data are collected, either by using sensors or by using social experiments, including showing videos of real and fake smiles and using human judgement to initiate statistical analysis.

In a sensor-based system, data are collected through using facial electromyography (EMG) [76-78]. This is a diagnostic technique used for recording facial muscle activity by placing electrodes on the face [79]. EMG has been widely used in measuring emotional reactions [80] and aids the diagnosis of facial nerve damage [81].

In [78], they used EMG to measure the “zygomaticus major”, which is a facial muscle that pulls the mouth angle to formulate the smile. This is done by placing electrodes on their subjects, thus allowing them to measure the smile intensity, duration and order of facial muscles movement or AU. Using this technique, they studied the difference between real and fake smiles; a smile formed with the help of the zygomatic major muscle (mouth corner raiser) and the orbicularis oculi muscle (cheek raises and forms crow’s feet around the eyes) and a non-Duchenne smile; a smile which only uses the zygomatic major muscle in the formation process. The result implies that spontaneous smiles not only formed faster than unspontaneous smiles but also used non-Duchenne muscle movements.

In [78] they used EMG to classify Duchenne and non-Duchenne smiles, identifying the muscles related to these smiles in the context of emotions (disgust, fear, anger, sadness, interest, surprise, pleasure). The result indicates that both

smiles have a high occurrence in the interest, surprise and pleasure expressions, however, the Duchenne smiles have significantly stronger EMG activity in the periocular and cheek muscle regions as compared to EMG activity in neutral faces.

2.2.2 Social Experimental Studies

From a social experimental perspective, a lot of research has been carried out on analysing people's reactions and influence on real and fake smiles. A research carried out by [82] shows individuals' preferences to work with individuals showing Duchenne (real) versus non-Duchenne (fake) smiles. In this experiment, participants performed two ostensibly unrelated tasks. First, they wrote essays about experiences of inclusion, exclusion or mundane events. Then, after successfully complete this task participants responded to 16 items assessing their levels of belonging, control, self-esteem and meaningful existence felt during the experience. Second, participants watched videos of individuals expressing Duchenne (10 videos) and non-Duchenne smiles (10 videos) and were asked to evaluate each individual as a potential partner for a project on which they might work. The results show participants have a greater preference to work with individuals displaying real smiles.

In [75], they investigated Chinese participants to judge for Duchenne and non-Duchenne smiles as real or fake smiles based on looking at mouth and eyes. In their experiment, 100 participants were asked to evaluate 20 videos and asked to rate each video as a real or fake smile. Afterwards, participants were asked to answer the question: "What part of the face was most useful for discriminating between fake and real smiles?" The results showed that participants highly

depend on information from the eyes to successfully distinguish between real and fake smiles and participants who choose eyes more accurately identify a Duchenne smile (real smile).

2.2.3 Computational Approaches

From a computing perspective, a lot of research has been using psychological studies as ground rules to distinguish between fake and real smiles using the machine. According to [78], a smile that raises the corners of the mouth and raises the cheeks with an appearance of eye wrinkles is considered as a Duchenne smile, whereas a smile that only raises the corners of the mouth is known to be a non-Duchenne smile. In terms of the FACS system, a Duchenne smile uses AU6 and 12 and a non-Duchenne smile uses only AU12 [74]. Moreover, accordioning to a set of research [83-86] they found that a Duchenne smile can be categorised as a true expression of emotion in comparison to a non-Duchenne smile.

In [87] they proposed a system to detect fake and real smiles based on detection of action units 6 and 12. This is done by using a Gabor filter and 2D principle component analysis using AdaBoost algorithm for features reduction. For classification, they used a support vector machine and gained 85.9% correct classification.

In [88] in their work, they divided the face into four regions (both eyes and mouth left/right). They applied simple principle component analysis (SPCA) to compute the eigenvector using grey scale image vector to compute a value cos ∅ and then used a neural network to classify the cos ∅ between fake and real smiles and gain 90% correct classification.

In [166] they propose deep convolution neural network for identifying fake and real smile by identifying different emotions percentage which signifies each type of the smile. They proposed a 9-layers CNN for detecting seven emotions (neutral, happy, disgust, fear, sad, anger and disgust) in the FERC-2013 dataset where they detect faces using viola jones algorithm and then resize the face low resolution (48*48 pixels) which these images will be used to train the CNN. Identifying the fake and real smile done by using the variations in the percentage of emotions which will be feed to specific software that discriminates each type of smile by the comparing the seven emotion percentages. For evaluating the proposed method, they use open source images and report very high accuracy on detecting a fake and real smile.

2.3 Gender Classification Using Smile Dynamics

Human face analysis can find its roots in various fields, including computer vision, physiology, biometrics, security and even medicine. Such work can provide a useful insight into an individual's health status[89], identity [90], beauty and behaviour, all of which depend on the non-invasive information obtained directly from the face.

Computer-based analysis of the human face can provide cues for personal attributes such as age, ethnicity and, more importantly, gender, in the present context. Gender classification, in this sense, can for example, aid as a biometric feature to improve its accuracy. Recent research into gender classification has faced challenging hurdles, mainly due to the reliance on static data in the form of facial images. Thus, texture-based facial image analysis has many inherent factors associated with image capture, such as lighting conditions, pose and

occlusions. In this regard, we have departed from texture-based analysis of facial images. Instead, we consider the analysis of the dynamic face, in particular, the dynamics of the smile, to be a superior alternative for gender classification.

Facial expressions are one of the prominent forms of visual and non-verbal communications for humans and are represented by facial muscle movements. Facial expressions are considered to be important indicators in understanding the emotional states of a person. Moreover, it is also plausible to assume that facial expressions provide an insight into the human mind, for example, the way one reacts to a specific situation can even inform us about the situation itself [91].

In this work, we study facial expressions in search of clues towards gender classification. This hypothesis has been supported by a set of cognitive physiological studies showing evidence of gender variances in facial expression, showing in [92-95]. In this section, we discuss recent research in smile and gender classification from two viewpoints, i.e. psychological and computational.

2.3.1 Psychological Perspectives

The psychological perspective dominates research using facial electromyography (EMG) [76-78], which is a diagnostic technique used for recording facial muscle activity by placing electrodes on the face [79]. EMG has been widely used in measuring emotional reactions [80] and contributing to diagnosis damages in facial nerves [81]. EMG has also been applied in gaming, where measuring a player's emotional state reflects the impact of games upon them. Furthermore, it had been applied in human computer interaction (HCI), where measuring the emotional state of users enhances the interaction between human and machine [96].

In [76] they used EMG for measuring angry and happy expressions. This is done by using EMG to measure corrugator and zygomatic muscle areas. The result shows that angry faces displayed increased in corrugator muscle activity whereas happy faces displayed increased zygomatic muscle activity. Additionally, these effects were more noticeable for females, mostly for the reaction to happy faces which conclude that females are more facially reactive than males.

In [78] they used EMG to classify between the Duchenne and non-Duchenne smiles, identifying the muscles related to these smiles in the context of emotions (disgust, fear, anger, sadness, interest, surprise, pleasure). As a result, they found that both smiles have a high occurrence in interest, surprise and pleasure expressions. Additionally, the Duchenne smiles have significantly stronger EMG activity in the periocular and cheek muscle regions as compared to EMG activity in the neutral faces. Moreover, a lot of psychological studies show a difference in human expressions [97] and show females express emotions more than males in terms of the smile. In fact, it has been documented that females smile more than males in a variety of social contexts [92]. Furthermore, the work in [98] suggests that on average females have more expressive smiles compared to males.

2.3.2 Computational Perspectives

From the computer vision perspective, gender classification can be divided into three main categories based on the feature extraction algorithm used. Feature extraction uses key points (distinguishing features) to represent objects. In gender recognition, feature extraction is the process of analysing the face to identify gender using the following models:

1. Geometric Model
2. Appearance Model
3. Hybrid Model (Appearance and Geometric).

The Geometric Model uses facial geometry as features, including the distances and size of facial features (eyebrows, eyes, nose and mouth). Furthermore, this model measures the movement of points through specific areas; these points are facial landmarks which describe the shape of the face and its components, such as the mouth or eyebrows. In [99] they use geometric features obtained from Canny edge detection to measure global features. The global features consist of interocular distance, the distance between the lips and the nose tip, the distance between the nose tip and the line joining the two eyes. The classification was computed by using a threshold value to distinguish between genders. In [100] they used principle component analysis (PCA) in order to compute eigenfaces. PCA is an algorithm used in an unsupervised technique for dimension reduction and data compression; they used the Minimum Distance Classifier for classification. In [101] they used 49 facial landmarks, produced by the cascade of linear regression, and tracked them using sparse optical flow, which was used to measure 27 geometric distances. For classification, they used a pattern classifier on labelled data with a support vector machine (SVM).

The Appearance Model can also be referred to as template matching or using an exemplar of the object. Appearance Model implies that objects look different under changes in lighting, colour, direction and at different scales. Furthermore, it describes the texture of the facial features. In [102] they used the Gabor function to extract the facial features texture. The Gabor function uses a set of wavelets with specific orientations and directions to represent different

textures. It is mathematically heavy and so is not applicable for real-time applications. Therefore, they used PCA to minimise the features extracted and reduce the processing time.

In [103] they used both the shape and texture information of the facial features. This is done first by dividing the face into smaller regions and then applying local binary pattern (LBP) histograms which have been formulated into a single vector. For classification, they use support vector machines (SVMs).

The Hybrid Model uses both the Appearance and Geometric Models. In [104] they used Discrete Cosine Transform (DCT) and Local Binary Pattern (LBP) algorithms as Appearance Models and Geometrical Distance Feature (GDF) for the Geometric Model. In [105] they divided the face into facial regions and applied 2DPCA into these regions and finally classified using SVM. In [106] they used HAAR wavelets to extract texture feature and Active Appearance Model to detect facial landmarks (83 landmarks). For classification, they used a support vector machine and radial basis function (RBF). In [107] they used a Local Directional Pattern (LDP) to identify appearance features. They used three steps to represent the face using LDP: first, LDP applied on the face image. Second, a histogram is extracted from each local region of LDP image. Third, building global representation using all the histograms are concatenated into one feature vector. For classification, they used a support vector machine (SVM).

In [90] they proposed an algorithm to identify gender based on the smile. Their proposed method used a set of algorithms to analyse the motion and texture of a smile to find a gender difference. They used a dense trajectories algorithm where it was used to extract a feature's point and track it using dense optical flow. Additionally, they used a Histogram of Oriented Gradients (HOG) and Video

Covariance Matrix Logarithm (VCML) to analyse appearance features and they used a Histogram of Optical Flow (HOF) and Motion Boundary Histogram (MBH) to analyse motion. All VCML, HOG, HOF and MBH are computed around a trajectory to identify smile characteristics. Finally, they used a Fisher vector to encode smiles; they applied first and second order statistics for each of the features produced by VCML, HOG, HOF and MBH. As a result, they show the proposed system is exceeding some state-of-the-art algorithms and the proposed system gains an average of 86% on the UvA-NEMO Smile Dataset.

2.4 Emotional Biometrics

Biometric techniques have been developed as capable techniques for recognising individuals and authenticating people. Usually, traditional authentication techniques can be manipulated, such as passwords, PINs and ID cards. Nevertheless, an individual's biological characteristics cannot be manipulated [108, 109], such as physiological characteristics including face, fingerprints, iris, ear and more; behavioural characteristics include gait (waking), signature, keystroke dynamics and facial expressions. In this section, we discuss recent work in different biometric research in the face and facial expressions.

2.4.1 Face Biometrics

Face biometrics has several advantages over other biometric methods, where an involuntary action is required (for example, fingerprinting or hand geometry detection done by placing the hand on a hand-rest in front of a camera as well as identifying the iris or retina). Furthermore, it can be done inactively without any noticeable action from the user's perspective, since face images can

be acquired from a distance by a camera. This feature gains a beneficial advantage for security and surveillance purposes.

In research by[110], they used an image processing method to extract 16 facial parameters, which include the ratio of distance, area and angles. For classification, they used Euclidean distance and gained 75% correct classification in a database of 20 different people using two images per person. In Brunelli and Poggio [111], based on[110] approach, they computed a 32 vector of geometric features which include nose width, length, mouth position and chin shape as shown in Figure 2-9. The second feature used grey-level template matching. As a result, they obtained a 90% recognition rate from a database of 47 people (four images per person) on geometric features and 100% recognition on template matching.

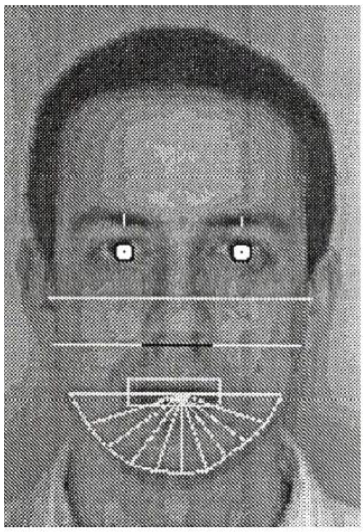


Figure 2-9: Geometrical features used by [103].

In [112] , they manually identified 35 geometric facial features as shown in Figure 2-10 and converted them into a 30-dimensional feature vector. For classification, they compared two algorithms: the nearest neighbour using

Euclidean distance and a new distance function for pattern recognition based on local second order statistics as estimated by modelling the training data as a mixture of normal densities. The result indicated 95% in a database of 685 images (a single image for each subject) compared with 84% from the nearest neighbour.

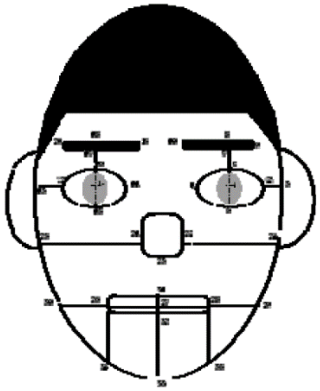


Figure 2-10: Manually identified 35 geometric facial features [104].

One of the earliest researches on face recognition using the Gabor function was presented by [113], where they presented an elastic bunch graph matching method. In this technique, they manually identified a set of fiducial points on the face and each fiducial point was a part of a fully connected graph where each fiducial point was considered as a node. Furthermore, they applied Gabor filters to a window around the fiducial point and identified an “arch” which is the distance between fiducial points which can be presented as a graph. combined these features graph into a stack-like structure introduce the “face bunch graph”. As a sequence, face bunch graph can be used to generate automatically by new face images by using elastic bunch graph matching.

For classification, they compared a face image graph with the pre-trained face images and the face with the highest similarity was considered to be a hit.

As a result, they gained 99% correct classification in a dataset containing 250 subjects. An update of this work was presented by [114] , using parametric models which eliminate the need to do the graph placement manually in work. This method worked on grey-level images, which have localised features, and it determined 16 facial fiducial points. As a result, they reported the same level classification as [112, 113].

In [115] they used PCA for face recognition. PCA is a statistical approach used for reducing the number of variables. using a Batch-2007 database to train PCA and produce eigenfaces and eigenvectors. For classification, they used the Euclidean distance between the test image and the eigenfaces. They didn't report the correct classification rate but stated a very accurate detecting rate.

In [116] they used PCA to create an attendance system. They started with face detection using the Viola Jones algorithm then cropped the face image and converted it to a vector using PCA. Furthermore, they normalised the face vector and projected it in the eigenface space. For classification, they computed the weight of the input image and compared it with pre-trained data weight using Euclidean distance. The result indicated a 0.099% error rate.

In [117] they proposed a geometric approach for face detection. The proposed system calculated geometric measurements such as area feature (eye, nose, mouth area), directional information of facial features' edge (horizontal, vertical, diagonal) and multiscale information using whole face information. Their proposed system can be summarised into five steps: (1) convert the grey image to black and white; (2) facial regions segmentation (eyes, nose and mouth); (3) calculate facial feature areas; (4) compute directional information using fast wavelet transform; (5) Selesnic's toolbox was used to compute a two-level Dual

Tree Discrete Wavelet Transform (RDT-DWT) for apprehension of the variations of pose, expression and orientations. As a result, they reported a 95% correct classification.

In [118] they used the texture analysis approach to face recognition. This was done by using Local Binary Patterns (LBP). LBP is an algorithm used to perform texture descriptors and it has been widely used in various applications. Using LBP on the face is derived by the fact that faces are a composition of micropatterns which can be described and analysed by LBP. In their work, they divided the face into regions of interest, where they used different kernel sizes, as shown in Figure 2-11, to identify a face scale differently and apply LBP. Using this approach, applying LBP in every single region identifies the facial feature related to that region and the grouping of all regions describes the texture and global geometry of the face. As result, they showed that the proposed method gained 97% correct classification on a FERET dataset.

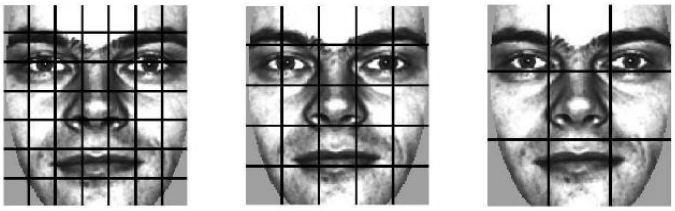


Figure 2-11: Facial image divided into rectangular regions.

In [119] they used a neural network to approach face recognition. A neural-based algorithm was presented to detect frontal views of faces. To reduce the dimensionality of face image they used PCA as a pre-step to train the Back Propagation Neural Network (BPNN). Using the Yale dataset, they obtained a

90% correct classification rate with lower execution time compared to the eigenfaces approach.

In [120] they use a non-statistics-based face representation approach and a local Gabor binary pattern histogram sequence (LGBPHS). In their approach, they segmented the face into regions and applied a Gabor filter, which created Gabor Magnitude Pictures (GMP), then using LBP they created a local Gabor Binary Pattern (LGBP) map. They applied their approach on the FERET dataset and used K-nearest neighbour (KNN) for classification. As a result, they gained 98% correct classification.

In [121] work, they used a hybrid model of machine learning and geometric features. The proposed system detects faces using the Viola Jones algorithm and detected 33 facial landmarks using an annotation process provided by [122] to describe geometric features. Moreover, they used a minimum redundancy maximum relevance (mRmR) algorithm in order to remove features' redundancy and identify features' relevance. For classification, they used a multiclass support vector machine (SVM) and they divided the dataset into three percentages: 30% for training, 20% for validation and 50% for testing. As a result, they gained 89% on the ORL dataset and 92% on the Caltech-Faces dataset.

Another type of biometric is gait, which is a research that tries to identify humans by the way they walk without taking into consideration any clothes, ethnicity or their background. As face recognition, gait has the advantage of not needing the subject's contact. Furthermore, gait biometric can be used in cases where finger and face are not applicable, for example, bank theft where thieves usually put on helmets, hoods, masks, glasses and gloves which make it hard to identify them using other known biometrics. On the other hand, there are a lot of

factors that affect gait recognition such as drunkenness, pregnancy and feet or joint injuries which can affect individuals' motion. According to [123] these factors have a similar effect in value to factors affecting other biometrics.

2.4.2 Facial Expressions based Biometric

Facial expression has been studied for further cues of mental health, longevity, emotions, gender and more. In terms of using it as biometric, the research is very limited. Facial expression biometric (FEB) can be identified as the way a person behaves and expresses emotions can indicate a person's identity [124]. Although the research of FEB is very limited there are a couple of psychological and computational studies that have tried to prove the existence of FEB. The majority of these researches study the facial expression biometric in terms of smile expressions. This is done by either using facial electromyographic (EMG), psychological studies, image processing or a hybrid model of them.

In [125] they represented a study of the stability of smile facial expressions from a psychological point view. Using facial electromyographic (EMG) to measure action unit 12 (zygomatic major muscles) associated with smile expression. AU12 showed a stable variation through two-year studies of the same individual. Furthermore, it reported recognised individuals based on their EMG reading for facial action units far above chance.

In [126] they studied the stability of smile expressions through time. In their research, they studied action units 6 and 12 to analyse 195 smiles from 95 individuals. Using automated facial software analysis and EMG (Figure 2-17) to analyse spontaneous smiles, the study covered two sessions with a year's

interval. As a result, facial EMG provided evidence to support smile expressions' stability over time.

Figure 2-12: Subject with EMG showing: neutral (a), AU6(b), AU6 + AU12(c).

In [125] he studied the individual differences in facial expression, stability over time and the ability to identify the person. His research included two studies. The first study used a 12-month interval for 65 adults showing a positive effective reaction to films. The positive effect was computed by measuring a zygomatic major muscle (main muscle involved in smiling) using facial electromyographic (EMG) and feature-point tracking using optical flow. The second study was carried out over four months including 85 middle-aged to older adults in a clinical interview, using the FACS system to manually code facial expression using pattern recognition to validate the value of facial behaviour signatures. As a result, both studies show the stability of facial expression over time and reports recognised individuals far above chance based on facial action units using EMG and feature-point tracking using optical flow.

One of the recent researches on FEB used facial tracker displacement [124]. Using active appearance models (AAMs) to identify facial landmarks, as

shown in Figure 2-18 to compute displacements, they used two images of a person: first, neutral emotion expression. Second, the peak of emotion which illustrates the movement of the face and not the face geometry. Using PCA and Euclidean distance on classification on the CK+ and Big Brother 3 datasets they showed statistical testing performed showed that such features can be used for personal identification.

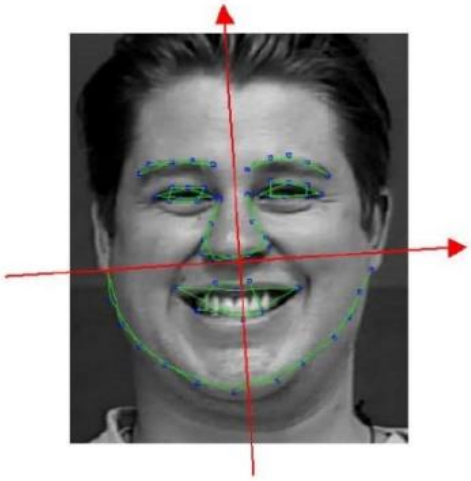


Figure 2-13: Tracker points.

2.5 Summary

In detecting facial expressions and action units, current research can be categorised into 4 approaches. They are appearance based, geometric based, hybrid models and CNN or DNN based approach. Appearance based approach tries to analyse the facial feature appearances. Geometric approach tries to identify the geometry of the face. The hybrid model uses both appearance and geometric approach to gain the benefit of both geometric and appearance based models. Finally, recent research carried on facial expressions use the convolutional neural network (CNN) or deep neural network (DNN) which is a

machine learning technique that uses a large number of images to train the machine to identify or classify an object.

According to recent research, both CNN and DNN have the highest classification rate followed by the hybrid, appearance and geometric approach. Although these approaches approve high detection rate to identifying the facial expressions we feel a new technique could be introduced to the field which is the motion model. In chapter 4, we propose an automated system to detect facial expressions based on the facial muscle movement.

For identifying posed and genuine smiles, we look at recent research in three different fields namely physiological, social and computational studies. Physiological studies use the EMG to measure the muscle movement through the nerve system in each type of the smile. Their results imply that real smiles have stronger EMG reading comparing to the posed smiles. From a social experimental point of view, they study people's reaction and judgment on identifying posed and genuine smiles through a series of experiments which include answering a questionnaire based on showing videos. The results indicate people use the area around the eyes to distinguish between the two types of smiles. From a computing perspective, they use the appearance, geometric and hybrid models to identify the difference between posed and genuine smiles.

Although the variety of research carried on posed and genuine smiles are promising, we study the possibility of imitating the physiological studies using image processing techniques which has not been addressed before. Moreover, we study the relationship between the facial features and their weight in formulating each type of smile which is illustrated in Chapter 5.

We study gender classification from two different viewpoints, i.e. physiological and computing perspectives. From a physiological perspective, where they use the EMG to measure the facial expressions for both genders. In terms of the smile expression, they found that female subjects have a more expressive smile than a male subject. From a computing perspective, gender classification has been approached mainly by using 3 models: appearance, geometric and hybrid model. All these models reported a very high classification rate when applied on faces with a neutral expression. Inspired by the finding of such physiological studies we study the intensity of smile expression of both genders using only the dynamic feature of the smile as the appearance model is considered to be biased. We discuss our method in detail in Chapter 6.

The present research on emotional biometrics is very limited and we approached this problem from two perspectives, namely physiological and computational. Physiological studies used EMG to study the stability of human emotion over time. Furthermore, they study the possibility of using EMG to determine human identity. Their results indicate that facial expressions are stabile over time and reported identifying individuals far above chance using the EMG. From a computing perspective, they use facial landmark displacement in a small dataset and reported that the identify individual can be recognised with above chance accuracy. Based on these findings, we create a system to find the possibility of using the dynamic characteristics of the smile as a biometric. We discuss this in detail in Chapter 7.

3 Enhancing Facial Feature Detection

Recent evolution in computer technologies has paved the way for the use of machinery through which human life can be influenced and enhanced by artificial intelligence. This encourages the need to develop a machine intelligence and “human-like” machines [127]. Computer vision, for example, is one of the largest examples of machine learning which tries to reproduce the human ability to visualise and analyse objects. Computer vision systems have been used widely for industrial purposes, surveillance and security [128], medical diagnosis and more. Furthermore, computer vision is moving towards generalised applications such as face detection [129], face recognition [129] and video coding techniques [130].

Humans can detect the face naturally in many environments with different lighting conditions. However, in terms of computer vision, this task is not easy. Human faces are complex objects that come in a variety of shapes, colours and poses [131] which make face detection one of the key problems in computer vision. Through face detection, one can identify emotions, poses and gestures. However, such applications require additional facial feature detection and tracking [50].

Face detection has been approached in many ways using different algorithms where each algorithm has its own strengths and weaknesses. For example, in [132] Singh and Chauhan used the skin colour to detect the face. This method starts with the assumption that only faces appear in the image. In [133] Pai and Ruan enhanced the work presented in [7] and added a “low pass filter” in order for better detection of skin under different lighting conditions. The

assumption made here is that the skin-like objects are not appearing in the image. In [134] Rowley and Baluja used a neural network to detect frontal faces, by using a sliding window over the image and applying a neural network to decide whether or not the window contains a face. In [135] Osuna and Freund used a support vector machine (SVM), which is a machine-learning technique that uses large data to identify objects. SVM detects faces by scanning the image for face-like patterns at different scales in order to cover possible scales of the face and uses SVM as its core classification algorithm to determine the face and non-face objects. In [136] Ping and Weng used template matching on multiresolution images to detect face templates and to enhance their detection. They added colour segmentation to identify the face and non-face classes using different types of colour spaces like CIEXYZ, YCbCr, YUV and YIQ. In [24] Viola-Jones identified the face by applying different types of algorithms (e.g. integral images, Haar-like features, Cascade filter and AdaBoost) that all enable the face to be found (see Section 2. Thus, there are a lot of different methods and algorithms that try to detect faces from images.

In addition to face detection, many applications need to identify and track facial features like eyebrows, eyes, nose and mouth. Such applications include face recognition, video processing and more as discussed in [50]. In [137] they presented the active shape model (ASM), which is a statistical model of shapes that anticipates the appearance of the objects in new images. ASM identifies facial features by using training datasets which start by identifying the facial feature counters manually which then use principle component analysis (PCA) to identify the variations in the training dataset; this process is called “shaped model”. In [138] they enhanced ASM and the present active appearance model

(AAM). AAM represents shapes statically based on the grey-level appearance of the object that can indicate different samples of the objects in new images or different conditions. Both AAM and ASM use PCA to identify the variations in the training dataset but the main difference between them is that ASM seeks to identify the shape of facial features by points constructing statistical models of them. On the other hand, AAM seeks to match the position model and represents the facial feature texture in the image [139]. In [68] they tried to identify facial features by identifying the texture through applying the Gabor function, which consists of edge detection, by applying Gabor wavelets on the image. In [140] and [141] they tried to apply a set of edge detection algorithms (e.g. Sobel, Prewitt, Roberts and Canny) to identify the facial feature boundaries. In [24] Viola-Jones used Haar-like features to identify the location of facial features by using a training dataset to identify the eyes, mouth and nose location.

In this chapter, we investigate the effect of localization search on cascade classifier used to detect facial features location. This done by dividing the face into regions of interest (ROI), where ROI contain areas with prominent facial features (eyes, eyebrows, nose and mouth). We use the Viola-Jones algorithm as an experiential environment to instigate the effect of ROI on the performance and accuracy of the algorithm. As the Viola-Jones algorithm proved its accuracy and performance in a real-time application in different environments and had been used in many different types of research [68, 142]. Furthermore, it has been applied to identify facial feature locations.

The application of this research can be beneficial in detecting facial feature location more accurately and effectively which can be used in a variety of application such as face analysis, emotions detection, face verification and more.

This work has been published in 16th UK Workshop on Computational Intelligence [167].

This chapter is structured as follows. Section 3.1 describes and explains the Viola-Jones algorithm and discusses the disadvantages of detecting facial feature location; Section 3.3: proposed method; Section 3.5: shows the results.

3.1 Viola-Jones Algorithm

In 2001, Viola-Jones presented a framework to detect objects using four key concepts: Haar-like features, integral images, AdaBoost and Cascade filters. The Viola-Jones algorithm has been widely used to detect faces and prove its accuracy in detecting front faces in images as well as in videos with different lighting conditions and in real time. The Viola-Jones algorithm consists of four key concepts [24] described in details in Appendix C.

3.2 Disadvantage of Facial Feature Detection using Viola-Jones Algorithm

Although the Viola-Jones algorithm is used to detect the front face, it has also been used widely in object detection and identifies facial feature location on the face. One of the disadvantages of this algorithm is that it has a high false positive percentage when applied on the face region to identify the facial feature location. Herein, and throughout the thesis, the false positive percentage (FPP), for a given type of facial feature, is defined as the ratio of the total number of facial features of the given type identified by the algorithm to the true number of facial features of the given type per image, expressed as a percentage. This disadvantage was determined by running a cascade classifier for each facial feature on the CK+ database [42] and identify the number of detected facial

feature for the corresponded cascade classifier. The CK+ database contains 123 front face people expressing one or more of the seven emotions (happiness, surprise, sadness, anger, fear, disgust and contempt). Additionally, it contains a set of image sequences that represent posed and non-posed (spontaneous) expressions that's start from natural state and reaches the peak of the emotion. A total of 831(123(subjects)*7(emotions)) image sequences were used to evaluate each facial feature cascade classier and determine its accuracy.

Figure 3-1 shows the number of faces detected to the ratio of each eyes and mouth. As shown, face to eye ratio average is (1: 1.813149665) and for the mouth (1: 4.476814593) which means that for each face detection there will be double (one face four eyes) and four times for the number of mouths (one face four mouths).

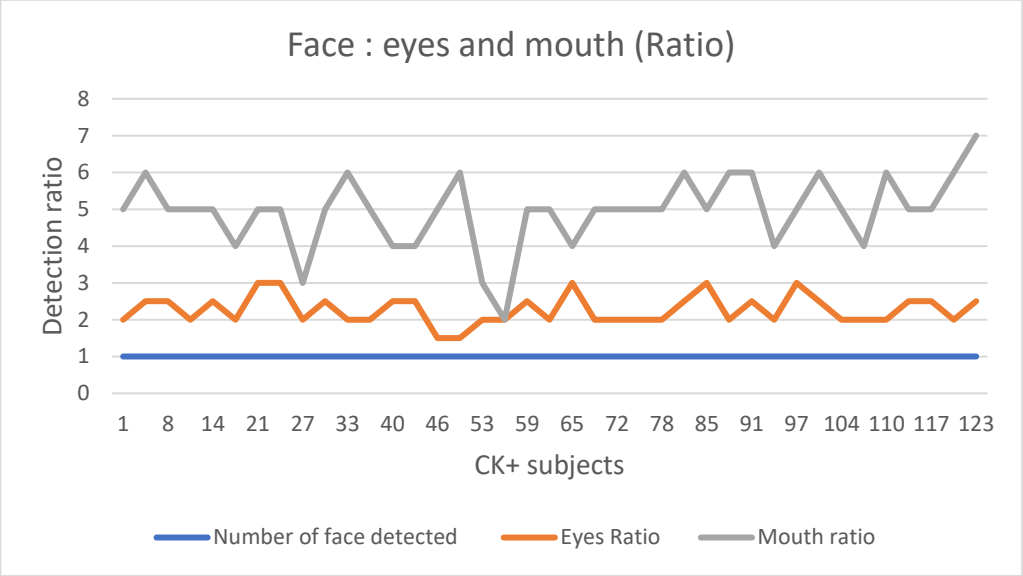


Figure 3-1: Face to mouth and eye ratio in Viola-Jones algorithm.

Figure 3-2 shows the actual number of left eyes and right eyes detected to the number of faces detected in the in each subject in the CK+ dataset.

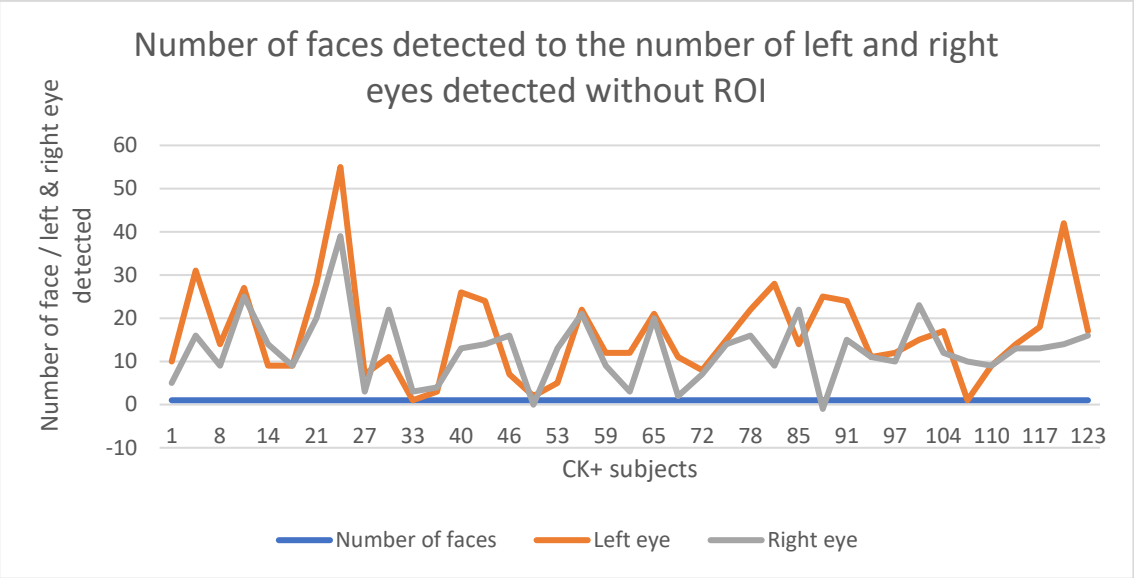


Figure 3-2: Number of faces detected to the number of left and right eyes detected.

3.3 Proposed Method

In order to minimise the false positive percentage for running the Viola-Jones framework, we defined a region of interest (ROI). ROI is an approximation for the location of facial features. Our approach was to divide the face into four main sections, where each section corresponded to an approximate area that contains one of the facial features (eyes, nose and mouth). Figure 3-3 illustrates this.

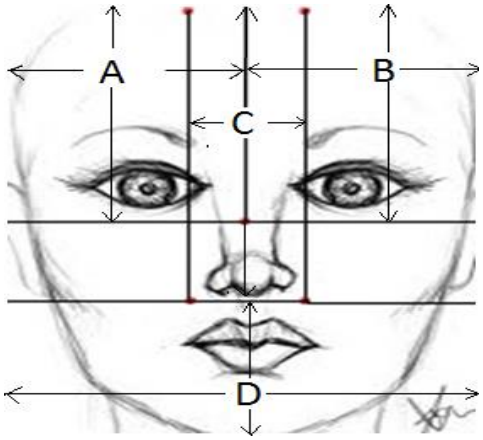


Figure 3-3: Facial features – four main parts.

As shown in Figure 3-3, part A includes the right eye, B includes the left eye, C includes the nose and, finally, D includes the mouth. For each area, we used the corresponding Viola-Jones Haar cascade where Viola-Jones trained three cascade filters, one for detecting both eyes and the other two to detect the left eye and right eye separately. We found that splitting the eye into two parts was better than looking for both eyes. This was because, in real-time tracking, if you lose track of one eye, you will lose track of both of them. Thus, by dividing the eyes into two areas we increase both the accuracy and eye tracking in real-time movement as losing one eye does not result in losing them both. Part C contains the nose area. By applying the nose cascade, we find that it has both a high false positive percentage and low accuracy and so by defining this area, we lower the false positive percentage and increase both the performance and the accuracy. Furthermore, we found that the nose area could be tracked even when we eventually lost track of the Haar cascade that could not define the nose in real time. Part D contains the mouth area.

Table 3-1 shows the start and end points for each part of the face. The face boundaries start from (X_s, Y_s) and the endpoints are ($X + w$, $Y + h$), where w is the width of the face and h is the height of the face.

Table 3-1 : Parts boundaries.

Face part	Facial Part	Start Point		End Point	
A	Right eye	X_s	Y_s	$(X + w) / 2$	$(Y + h) / 2$
B	Left eye	$(X + w) / 2$	Y_s	$(X + w)$	$(Y + h) / 2$
C	Nose	$(X + w) * 3/5$	Y_s	$(X + w) * 5/8$	$(Y + h) * 5/8$
D	Mouth	X_s	$(Y + h) * 3/5$	$(X + w)$	$(Y + h)$

3.4 Results

We tested our approach using the Cohn-Kanade Extended Facial Expression Database (CK+) [42]. The experiment included 123 people with front face pose. Figure 3-4 shows the detection rate for the three facial features (left eye, right eye and mouth) applying the Viola-Jones cascade filter without ROI showing a 90% false positive for left eye, right eye 73% and mouth 348%, which means that for each face detected there are four eyes and four mouths. Using ROI shows that the percentages of false positive for both eyes are 0% and mouth is dropped to 26% due to the weak classifier of mouth cascade.

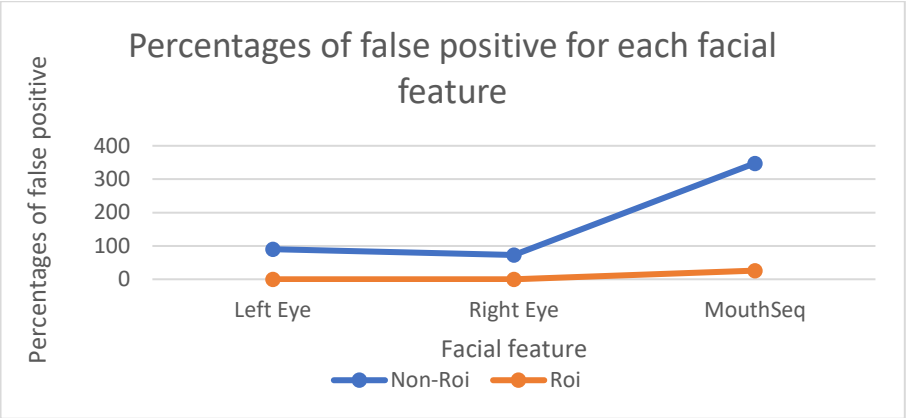


Figure 3-4 : Comparison between Viola-Jones with ROI and Viola-Jones without ROI.

Figure 3-5 demonstrates the number of face detections in the CK+ database and the number of left and right eyes detected. As a sample, subject 8 showed that it detected one face 3 left eyes and 2 right eyes, which showed the high percentage of false positive in the original algorithm.

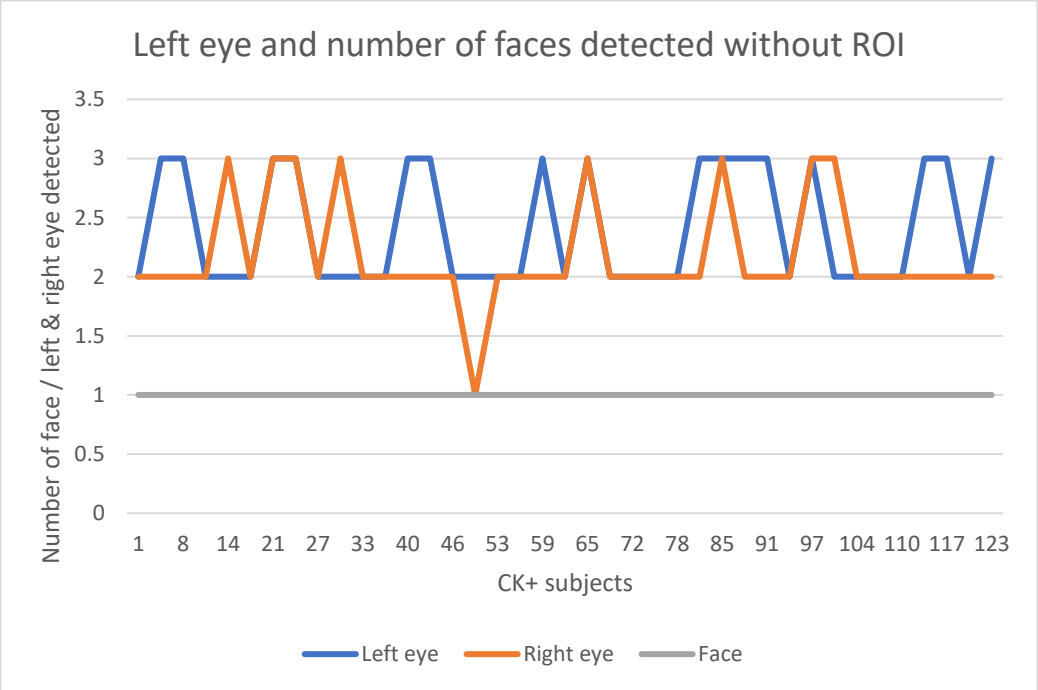


Figure 3-5 : Comparison between number of left and right eyes detected and number of faces without ROI.

Figure 3-6 demonstrates the number of face detections in the CK+ database and the mouth number detected. Due to the weak cascade classifier, the mouth detection has a high false positive percentage compared to the number of faces detected.

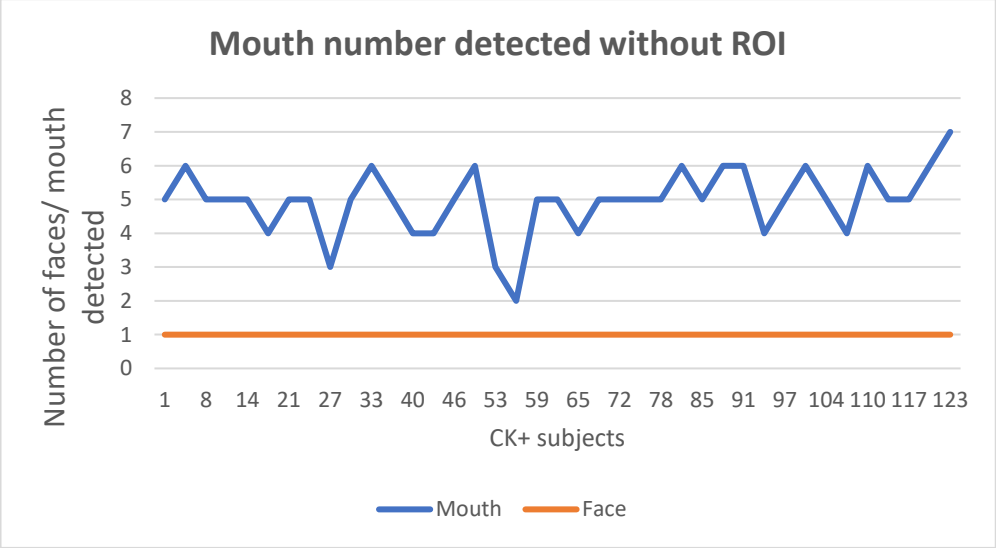


Figure 3-6 : Comparison between number of mouths detected and number of faces.

After applying ROI, we reduced the false positive percentage for each of the facial features. Figure 3-7 shows the number of faces detected in the database and the number of left and right eye detections. This figure shows that the number of left and right eyes is consistent with the number of faces detected.

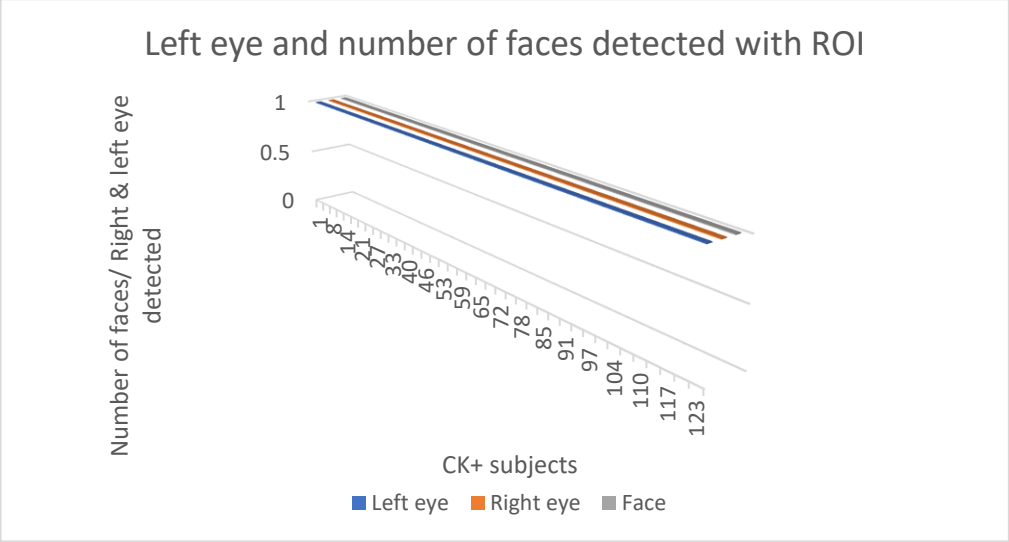


Figure 3-7 : Comparison between number of left and right eyes detected and number of faces with ROI.

Figure 3-8 shows mouth number with the number of faces detected using ROI, showing the number of mouths detected per face that are similar to the face number.

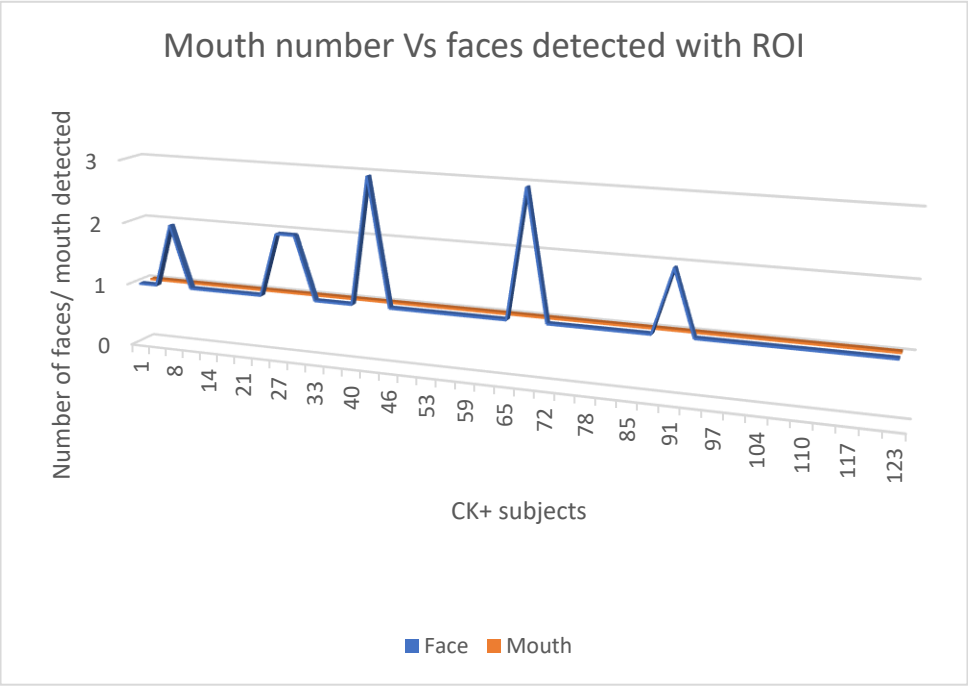


Figure 3-8 : Comparison between number of mouths detected and number of faces with ROI.

3.5 Conclusions

Using ROI enhances both the performance and accuracy of facial feature detection in the Viola-Jones algorithm by reducing the size and the location of each facial feature location. This results in a better detection rate and reduces the false positive percentage in the Viola-Jones algorithm. Moreover, the performance of the algorithm for real-time applications was enhanced since the area that needed to be examined was reduced.

Future work will investigate and analyse the effect of region of interest on other facial feature detection algorithms in terms of accuracy and performance. This approach can be used in identifying the facial feature location in order to analyse facial expressions which we describe in the next chapter.

4 An Automated System to Analyse and Detect Action Units and Facial Expressions

Humans interact more naturally with each other compared to interacting with machines. During face-to-face communication, people exchange information through verbal and non-verbal communication. According to different research [143], verbal communication contains one-third of human communication and non-verbal contains the other two-thirds. Verbal communication, such as a phone call or voice message, is often sufficient to communicate with other people. On the other hand, non-verbal communication includes visual cues such as body language, voice and physical appearance (facial expressions). Considering the emotional meaning, facial expressions are one of the main components of social communication. Facial expressions not only represent the emotional state but contain more information about human mood, relationship, health and the way human react to specific situations.

To approach more natural interaction with machines, different facial expressions should be understood to imitate the human way to communicate with each other. There are many definitions of facial expression but mainly it represents the emotional states characterised by facial muscle movement. Essentially, from an anatomical point of view, facial expressions are formed when a person experiences an emotion – different inner organs trigger facial muscles to change their structure and express emotions. The advantages of facial expression analysis include making more flexible and robust interactions between human and machine and enhancing user interface structure.

As previously discussed in Section 1.5, the facial action coding system (FACS) is one of the most well-known systems for analysing facial expressions. FACS uses a set of action units which represent muscle or a set of muscles movement to represent facial expressions. A total of 46 AU was produced by Ekman which can be used to construct or deconstruct one of the six basic emotions (surprise, happiness, fear, sadness, anger, disgust). Furthermore, Ekman presents the emotional facial action coding system (EMFACS) where he shows the construction of actions units related to each emotion, as shown in Section 1.5.1.

From a computing perspective, facial expressions analysis has been an active field for more than a decade. There are different approaches that try to analyse and detect facial expression and action units. Generally, recent research uses texture, geometric distance, machine learning or a hybrid system containing both texture and geometric distance. Recent research is discussed in detail in Section 2.1.

Although the high detection rate reported in the literature review we found that there is no research carried on identifying facial expression using facial muscle movement. Which shows how much facial expressions are encoded within the motion model. In this research, we propose a new methodology for facial action unit detection by analysing facial muscles movement only. This is done by constructing a connected web of regions of interest (ROI) that cover all possible movement of the facial features. Motion detection is done by applying dense optical flow in this ROI. Finally, we present a motion vector re-calculation engine (MVRE) which analyses these motions and converts them to AU units by using a rule-based system and motion profile technique.

The application of analysing and detecting facial expressions is massive where it can be used in analysing interviews, customer feedback regarding a certain product and check mental health through facial expressions. Furthermore, it can be used in social behaviour study where it will automate the evaluation process of human's emotions. Finally, it makes the machine more human where it will analysis and responding to emotions and interact with human using none-variable communications.

4.1 Methodology

Based on the previous work carried out on AU detection in Section 2.1, we designed an automated system to detect action unit activation based on facial muscles movement. Our system is an imitation of how facial muscles move to formulate action units.

We use an image sequence to analyse action units and facial expressions as it contains more information. A research by [21] compared detecting facial expression accuracy between static image (single frame) and image sequence. They found that a still image has less accuracy compared to a video sequence because a single image contains less information than a sequence of images.

Figure 4-1 shows the proposed framework for facial motion analysis. Our proposed framework contains three main phases: face detection, motion analysis and classification.

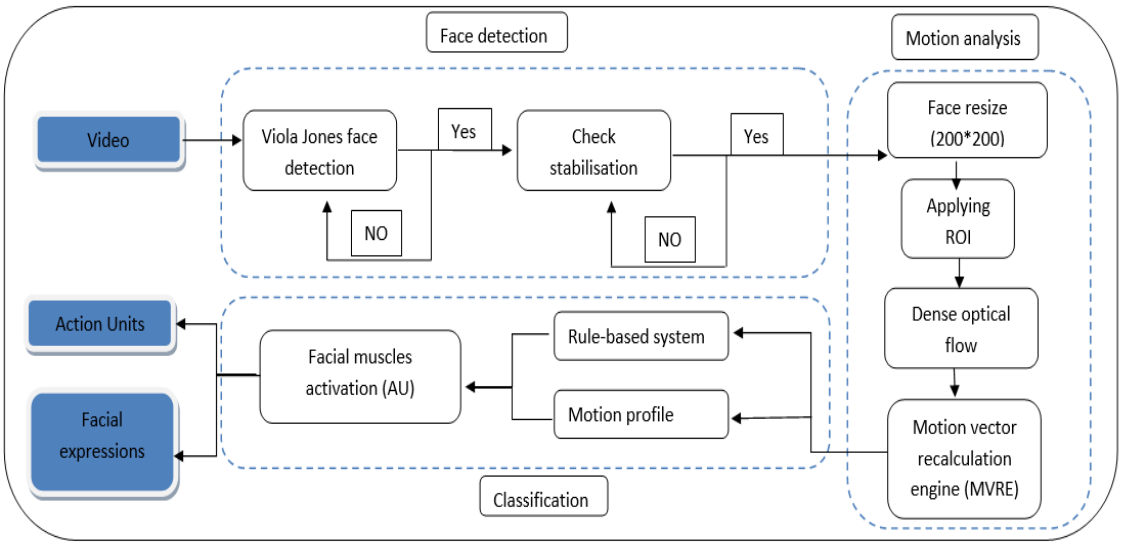


Figure 4-1 : Proposed framework.

The first phase is face detection which contains two main parts: face detection and stabilisation check. Detecting the face is done using the Viola-Jones algorithm. For more details refer to Section 3.1.

The second part is a stabilisation check, generally the Viola-Jones algorithm, which uses cascade classifiers (a strong classifier contracted by linear weighted weak classifiers). Detecting the face is done by combining these weak classifiers which produce a strong classifier. The combined computation of these weak classifiers will return slightly different face locations each time the algorithm is applied [144]. As a sequence, it will affect the motion analysis algorithm as it will be considered as a false movement. To solve this problem, we compute the facial normalisation parameter (FNP). FNP is used to stabilise the detection window in the Viola-Jones algorithm and can be computed using the following equation:

$$FNP = \left| Face\ Loc\ (i)_{(X,Y)} - Face\ Loc\ (i+1)_{(X,Y)} \right| , \quad (4.1)$$

$$FNP \leq €_1 , \quad (4.2)$$

where $Face\ Loc\ (i)_{(X,Y)}$ represents the face, location retrieved from the frame i and $Face\ Loc\ (i+1)_{(X,Y)}$ represents the face location retrieved from the next frame (i +1). FNP can be computed by subtracting the location of both faces and taking the absolute value. To detect stabile face location, we compare the FNP with the threshold value $€_1$. To identify the $€_1$ value, we create videos containing only the first frame for each subject to make sure the face has not moved to identify the $€_1$ value and stability of viola jones algorithm, a total of 861 videos were created from the CK+ dataset. By trying different $€_1$ value we determine to 10 pixels where it shows a table detection window.

4.1.1 Motion Analysis

The second phase is motion analysis which contains three main parts: applying region of interest (ROI), applying dense optical flow and Motion Vector Re-Calculation Engine (MVRE).

Applying ROI is done by first resizing the face for (200*200) to unify the allocation of each ROI. Second, we divide the face into three main regions: the upper, middle and lower areas. The upper area includes eyes, eyebrows and forehead, the middle area includes the cheeks and nose, and the lower area contains the mouth and chin. To capture different movements within each facial feature we apply a set of regions of interest (ROI) over each facial feature, where

ROI represents an approximate location for different parts of facial features, as shown in Figure 4-2.

Allocating ROI is built from two main concepts: facial muscle autonomy and facial electrograph (EMG) electrode location. First, we study the facial muscle autonomy and how the muscles are connected to each other. Generally, the adjustment facial muscle moves as one unit since they are connected and are responsible for the same facial muscle movement. Secondly, using EMG location, EMG consists of measuring muscle activity by electronic impulses that cause the muscle to move. Recent research[79, 145, 146] carried out on EMG defines the facial expression by plotting a set of electrodes to measure nerve system pulses; these pulses trigger muscles to move and formulate facial expression. Our main goal from this research is the EMG location of electrodes placed on the face. For further details refer to Appendix A.

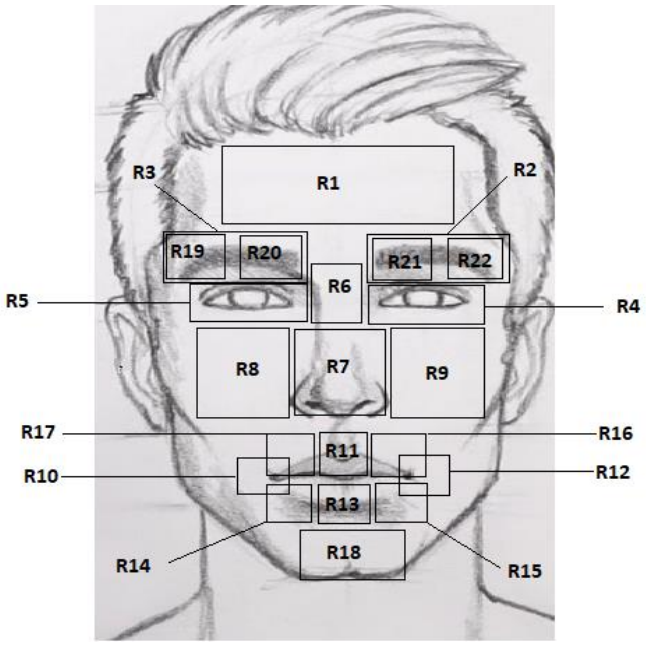


Figure 4-2: Regions of interest.

A total of 22 ROI are placed over each facial feature to cover all possible movements in that area. The distribution of facial features using ROI is as follows: the upper area contains nine ROI to measure the forehead (R1), eyebrows (R3, R2, R19, R20, R21, R22) and eyes (R5, R6). The middle area contains four ROI to measure the left cheek (R8) and right cheek (R9) and the nose (R7, R6). The lower area contains nine ROI to measure both mouth (R10 to R17) and chin (R18) movement.

To detect the motion in the face, we apply a dense optical flow algorithm by Gunnar Farneback [147]. An optical flow algorithm is usually applied to image sequences that have a small-time gap between them to identify pixel displacement or image motion. According to [49], optical flow works on two assumptions: the pixel intensities of an object do not change between consecutive frames and neighbouring pixels have a similar motion. In previous research [21, 57], optical flow was used to track points as “facial landmarks” in multiple images/videos and determine how these points moved. The output of this algorithm is a motion vector map representing the direction and displacement value of each pixel in the image. For more details refer to Appendix B.

As optical flow produces orientation and magnitude of each pixel in each ROI, we need to convert these motion vectors into action unit representation. We present the Motion Vector Re-Calculation Engine (MVRE), which is used to convert the motion vectors computed by optical flow, to features. These features are used to distinguish different action units (AU) which will be used in the classification phase. Furthermore, the MVRE filters error motion caused by different factors like optical flow equations or the environment.

4.1.2 MVRE

MVRE contains mainly from two parts: voting system and motion converter. The voting system represents the connection between each ROI. Furthermore, it is responsible for averaging the movement over the ROI and eliminating any anomaly caused by different factors like errors in computing the flow and environment conditions. The voting system consists of a web of connected ROI as shown in Figure 4-3. The voting system works on two adjustment levels excluding the eyes.

As an example, when a movement occurs in the front head area, the eyebrows area will be affected as well. As a sequence to check the correct movement occurs in ROI related to these areas (R1–R3, R19–R22) we calculate the overall movement and exclude any anomalies. As an example of this anomaly, the calculation is affected by environmental factors which will confuse the decision-making process of action unit activation.

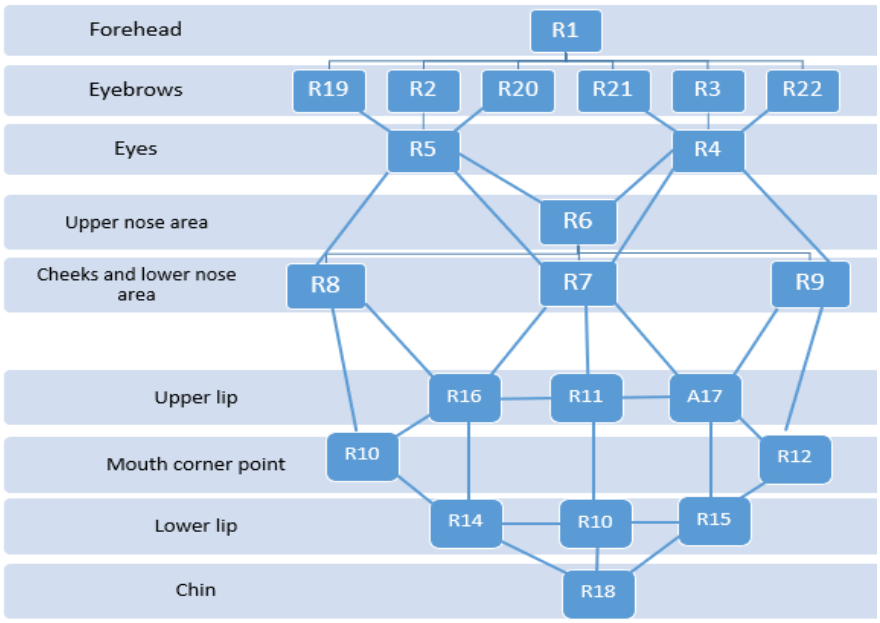


Figure 4-3: MVRE ROI connections.

The second part of MVRE is the motion converter which consists of a set of motion profile techniques that identify different movements in facial features. Forwarding the output of the voting system to motion converter ,Tables 4-1,4-2 ,4-3 shows a set of rules that represent the motion profile technique for different movements in eyes, mouth and cheek areas respectively. As an example of computing E1 table 4-1, we check the motion vector with the direction going up in the area R1–R3 and R19–R22 and using equation 4.3 we compute the total flow f in the related ROI R_i in a specific direction d where regions i are part of facial feature x.

$$MVRE\ code_{(x)} = \sum f\ (\overrightarrow{R_i^d})\quad where\ i \in x \tag{4.3}$$

These tables were constructed using the description of each AU and the corresponding ROI location. Section 1.5.1.1, Figure 1-7 contains AUs description and the corresponded muscle movement. As an example, E1 which represent the eyebrows moving up we measure the motion vector moving up in ROI (Figure 4-2: R1, R2, R3, R19, R20, R21, R22) These MVRE codes will be used later in creating the rules in the classification phase.

Table 4-1: MVRE codes for the eyes area.

Eye area movement	ROI movement vector	MVRE Code

Eyebrows moving up	$\overrightarrow{R_1^{Up}}, \overrightarrow{R_2^{Up}}, \overrightarrow{R_3^{Up}}, \overrightarrow{R_{19}^{Up}}$ $, \overrightarrow{R_{20}^{Up}}, \overrightarrow{R_{21}^{Up}}, \overrightarrow{R_{22}^{Up}}$	E1
Eyebrows moving down	$\overrightarrow{R_1^{Down}}, \overrightarrow{R_2^{Down}}, \overrightarrow{R_3^{Down}}, \overrightarrow{R_{19}^{Down}}$ $\overrightarrow{R_{20}^{Down}}, \overrightarrow{R_{21}^{Dowb}}, \overrightarrow{R_{22}^{Down}}$	E2
Eyebrows moving towards each other	$\overrightarrow{R_3^{Right}}, \overrightarrow{R_{19}^{Right}}, \overrightarrow{R_{20}^{Right}}$ $\overrightarrow{R_2^{Left}}, \overrightarrow{R_{21}^{Left}}, \overrightarrow{R_{22}^{Left}}$	E3
Eyebrows moving apart	$\overrightarrow{R_3^{Left}}, \overrightarrow{R_{19}^{Left}}, \overrightarrow{R_{20}^{Left}}$ $\overrightarrow{R_2^{Right}}, \overrightarrow{R_{21}^{Right}}, \overrightarrow{R_{22}^{Right}}$	E4
Eyebrows inner part moving up	$\overrightarrow{R_2^{Up}}, \overrightarrow{R_3^{Up}}, \overrightarrow{R_{19}^{Down/Nutral}}$ $\overrightarrow{R_{20}^{Up}}, \overrightarrow{R_{21}^{Up}}, \overrightarrow{R_{22}^{Down/Nutral}}$	E5
Eyebrows inner part moving down	$\overrightarrow{R_2^{Down}}, \overrightarrow{R_3^{Down}}, \overrightarrow{R_{19}^{Down/Nutral}}$ $\overrightarrow{R_{20}^{Down}}, \overrightarrow{R_{21}^{Down}}, \overrightarrow{R_{22}^{Down/Nutral}}$	E6
Eyebrows outer part moving up	$\overrightarrow{R_2^{Up}}, \overrightarrow{R_3^{Up}}, \overrightarrow{R_{19}^{Up}}$ $\overrightarrow{R_{20}^{Down/Nutral}}, \overrightarrow{R_{20}^{Down/Nutral}}$ $\overrightarrow{R_{22}^{Down}}$	E7
Eyebrows outer part moving down	$\overrightarrow{R_2^{Down}}, \overrightarrow{R_3^{Down}}, \overrightarrow{R_{19}^{Down}}$ $\overrightarrow{R_{20}^{Up/Nutral}}, \overrightarrow{R_{20}^{Up/Nutral}}$ $\overrightarrow{R_{22}^{Down}}$	E8
Eye closing	$\overrightarrow{R_4^{Down}}, \overrightarrow{R_5^{Down}}$	E9
Eye opening	$\overrightarrow{R_4^{Up}}, \overrightarrow{R_5^{Up}}$	E10

Table 4-2 : MVRE codes for the mouth area.

Mouth area movement	**ROI movement vector**	**MVRE Code**
Mouth narrow	$\overrightarrow{R_{10}^{Left}}, \overrightarrow{R_{14}^{Left}}, \overrightarrow{R_{16}^{Left}}$ $\overrightarrow{R_{12}^{Right}}, \overrightarrow{R_{15}^{Right}}, \overrightarrow{R_{17}^{Right}}$	M1

Mouth stretch	$\overrightarrow{R_{10}^{Right}}, \overrightarrow{R_{14}^{Right}}, \overrightarrow{R_{16}^{Right}}$ $\overrightarrow{R_{12}^{Left}}, \overrightarrow{R_{15}^{Left}}, \overrightarrow{R_{17}^{Left}}$	M2
Mouth corner moving apart	$\overrightarrow{R_{10}^{Right}}, \overrightarrow{R_{14}^{Right}}, \overrightarrow{R_{16}^{Right}}$ $\overrightarrow{R_{12}^{Left}}, \overrightarrow{R_{15}^{Left}}, \overrightarrow{R_{17}^{Left}}$	M3
Mouth corners moving toward each other	$\overrightarrow{R_{10}^{Left}}, \overrightarrow{R_{14}^{Left}}, \overrightarrow{R_{16}^{Left}}$ $\overrightarrow{R_{12}^{Right}}, \overrightarrow{R_{15}^{Right}}, \overrightarrow{R_{17}^{Right}}$	M4
Mouth corner moving down	$\overrightarrow{R_{10}^{Right}}, \overrightarrow{R_{14}^{Right}}, \overrightarrow{R_{16}^{Right}}$ $\overrightarrow{R_{12}^{Left}}, \overrightarrow{R_{15}^{Left}}, \overrightarrow{R_{17}^{Left}}$	M5
Mouth corner moving up	$\overrightarrow{R_{10}^{Right}}, \overrightarrow{R_{14}^{Right}}, \overrightarrow{R_{16}^{Right}}$ $\overrightarrow{R_{12}^{Left}}, \overrightarrow{R_{15}^{Left}}, \overrightarrow{R_{17}^{Left}}$	M6
Upper lip moving up	$\overrightarrow{R_{7}^{Up}}, \overrightarrow{R_{11}^{Up}}, \overrightarrow{R_{16}^{Up}}, \overrightarrow{R_{17}^{Up}}$	M7
Upper lip moving down	$\overrightarrow{R_{7}^{Down}}, \overrightarrow{R_{11}^{Down}}, \overrightarrow{R_{16}^{Down}}, \overrightarrow{R_{17}^{Down}}$	M8
Lower lip moving up	$\overrightarrow{R_{13}^{Up}}, \overrightarrow{R_{14}^{Up}}, \overrightarrow{R_{15}^{Up}}, \overrightarrow{R_{18}^{Up}}$	M9
Lower lip moving down	$\overrightarrow{R_{13}^{Down}}, \overrightarrow{R_{14}^{Down}}, \overrightarrow{R_{15}^{Down}}, \overrightarrow{R_{18}^{Down}}$	M10
Jaw moving up	$\overrightarrow{R_{13}^{Up}}, \overrightarrow{R_{18}^{Up}}$	M11
Jaw moving down	$\overrightarrow{R_{13}^{Down}}, \overrightarrow{R_{18}^{Down}}$	M12
Mouth open (general movement)	$\overrightarrow{R_{7}^{Up}}, \overrightarrow{R_{11}^{Up}}, \overrightarrow{R_{16}^{Up}}, \overrightarrow{R_{17}^{Up}}$ $\overrightarrow{R_{13}^{Down}}, \overrightarrow{R_{14}^{Down}}, \overrightarrow{R_{15}^{Down}}, \overrightarrow{R_{18}^{Down}}$	M13
Mouth closing (general movement)	$\overrightarrow{R_{7}^{Down}}, \overrightarrow{R_{11}^{Down}}, \overrightarrow{R_{16}^{Down}}, \overrightarrow{R_{17}^{Down}}$ $\overrightarrow{R_{13}^{Up}}, \overrightarrow{R_{14}^{Up}}, \overrightarrow{R_{15}^{Up}}, \overrightarrow{R_{18}^{Up}}$	M14

Table 4-3 : MVRE codes for the cheeks.

Cheeks area movement	ROI movement vector	MVRE Code

Cheek moving up	$\overrightarrow{R_8^{Up}}, \overrightarrow{R_{10}^{Up}}, \overrightarrow{R_{14}^{Up}}, \overrightarrow{R_{16}^{Up}}$ $\overrightarrow{R_9^{Up}}, \overrightarrow{R_{12}^{Up}}, \overrightarrow{R_{15}^{Up}}, \overrightarrow{R_{17}^{Up}}, \overrightarrow{R_{24}^{Up}}$	C1
Cheek moving down	$\overrightarrow{R_8^{Down}}, \overrightarrow{R_{10}^{Down}}, \overrightarrow{R_{14}^{Down}}, \overrightarrow{R_{16}^{UpDown}}$ $\overrightarrow{R_9^{Down}}, \overrightarrow{R_{12}^{Down}}, \overrightarrow{R_{15}^{Down}}, \overrightarrow{R_{17}^{Down}}, \overrightarrow{R_{24}^{Down}}$	C2
Cheek moving apart	$\overrightarrow{R_8^{Left}}, \overrightarrow{R_{10}^{Left}}, \overrightarrow{R_{14}^{Left}}, \overrightarrow{R_{16}^{Left}}, \overrightarrow{R_{23}^{Left}}$ $\overrightarrow{R_9^{Right}}, \overrightarrow{R_{12}^{Right}}, \overrightarrow{R_{15}^{Right}}, \overrightarrow{R_{17}^{Right}}, \overrightarrow{R_{24}^{Right}}$	C3
Cheeks moving towards each other	$\overrightarrow{R_8^{Right}}, \overrightarrow{R_{10}^{Right}}, \overrightarrow{R_{14}^{Right}}, \overrightarrow{R_{16}^{Right}}, \overrightarrow{R_{23}^{Right}}$ $\overrightarrow{R_9^{Left}}, \overrightarrow{R_{12}^{Left}}, \overrightarrow{R_{15}^{Left}}, \overrightarrow{R_{17}^{Left}}, \overrightarrow{R_{24}^{Left}}$	C4

4.1.3 Classification

For classification, we use two approaches: a rule-based system and the motion profile. The rule-based system uses MVRE measurement as shown in Tables 4-1, 4-2,4-3 to classify one or more action units as shown in Tables 4-5, 4-4. Motion profile which is a general representation of how ROI movement when a certain action unit or action units activated. Motion profile was generated by monitoring subjects expressing different action units when expressing certain facial expressions. Basically, the motion profile uses one or more of the MVRE codes to detect multiple action activation.

These tables were constructed on two concepts: the description of the AUs and the visualization aspects of the FACS system. Description of the AUs is mention in Section 1.5.1.1 where each AU have the movement described by the related facial feature which was provided by the FACS system. The visualization aspect [148] contains a short video of the each AU shows the corresponded facial

feature muscle movement which we analysis to determine the movement in specific AU.

Table 4-4: MVRE classification rules.

Action units	**Description**	**Rules**
AU 1	**Inner eyebrows raised**	$E_5 > E_6$ $E_1 > E_2$ $E_3 > E_4$
AU 2	**Outer eyebrows raised**	$E_7 > E_8$ $E_1 > E_2$ $E_6 > E_5$
AU 4	**Eyebrows lowered**	$E_2 > E_1$
AU 5	**Eyebrows raised**	$E_1 > E_2$
AU 6	**Cheek raised**	$C_1 > C_2$ $C_3 > C_4$
AU 7	**Eye tightened**	$E_9 > E_{10}$ $E_2 > E_1$ $C_1 > C_2$
AU 9	**Nose wrinkle**	$C_1 > C_2$ $M_5 > M_6$ $E_2 > E_1$
AU 10	**Mouth upper lip raised**	$M_5 > M_6$ $M_5 > M_7$ $E_2 > E_1$
AU 12	**Mouth corner pulled**	$M_3 > M_4$ $C_1 > C_2$
AU 15	**Lip corner depressor**	$M_6 > M_5$ $M_7 > M_8$
AU 16	**Mouth lower lip depressor**	$M_8 > M_7$ $M_8 > M_5, M_8 > M_6$

Table 4-5: MVRE classification rules.

Action units	Description	Rules
AU 17	**Chin raised**	$M_9 > M_{10}$
AU 18	**Mouth lips puckered**	$C_1 > C_2$
AU 20	**Mouth lips stretched**	$M_2 > M_1$ $M_3 > M_4$
AU 22	**Mouth lips funnelled**	$M_{13} > M_{14}$ $M_2 > M_1$ $M_4 > M_3$ $M_{10} > M_9$
AU 23	**Mouth lips tightened**	$M_1 > M_2$ $M_8 > M_7$ $M_4 > M_3$ $M_9 > M_{10}$ $M_{11} > M_{12}$
AU 25	**Mouth lips a part**	$M_5 > M_6$ $M_7 > M_8$ $M_9 > M_{10}$
AU 26	**Jaw down**	$M_{10} > M_9$
AU 27	**Mouth stretch**	$M_2 > M_1$ $M_2 > M_4$

As an example of AU4 (lowering eyebrows) activation, the rule-based system uses MVRE equations shown in table 1 to check three rules (E1 > E2, E5 > E6, E5 > E7). Satisfying these rules will lead to activation of AU4. Furthermore, as an example of motion profile, Figure 4-4 shows activation of AU4 (lowering eyebrows) and AU12 (cheek raised). For illustrative purposes, flow value is set to 1 to identify full movement in a certain direction in the related ROI. This figure implies AU4 have general motion going down and AU12 have general motion (up, left) in the left cheek and (up, right) in the right cheek.

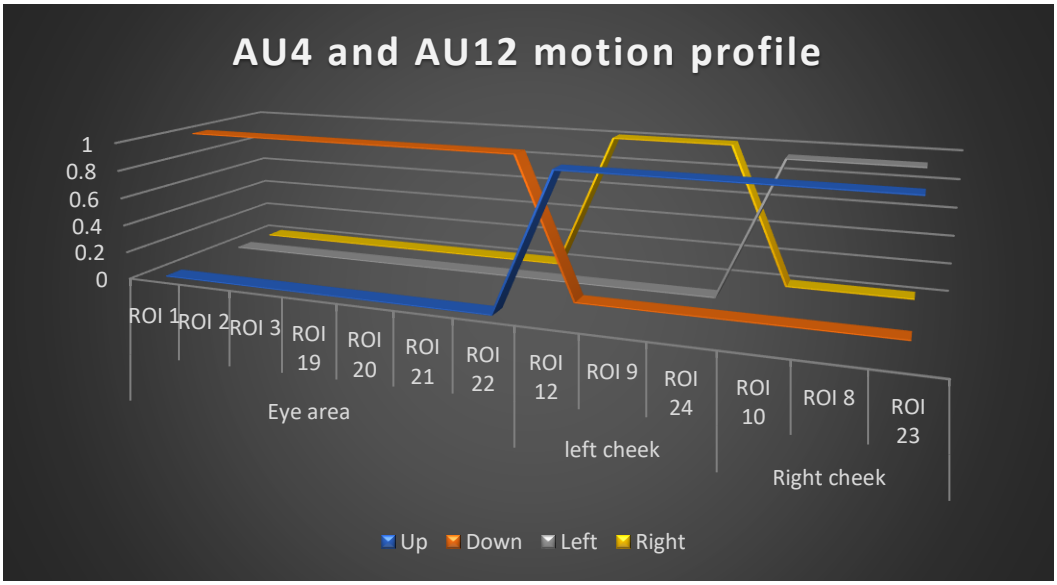


Figure 4-4: Motion profile for AU4 and AU12.

For classifying emotions, we use emotional FACS presented by Ekman to identify emotions using AU. EMFACS present the combination of AU to produce one of the six basic emotions as shown in Table 4-6:

Table 4-6: Emotional facial action coding system (EMFACS).

Emotions	Facial Action Units (FAU)
Happiness	AU 6 + AU 12
Surprise	AU 1 + AU 2 + AU 5 + AU 26
Sadness	AU 1 + AU 4 + AU 15
Fear	AU 1 + AU 2 + AU 4 + AU 5 + AU 7 + AU 20 + AU 26
Anger	AU 4 + AU 5 + AU 7 + AU 23
Disgust	AU 9 + AU 15 + AU 16

4.2 Implementations

To implement our approach, we use the OpenCV platform. OpenCV stands for Open Source Computer Vision Library, which is a software library that provides an infrastructure for computer vision applications. Furthermore, OpenCV is a free source (BSD license) [149] and its library contains more than 2,500 implemented and optimised algorithms, encompassing classic and state-of-the-art computer vision and machine-learning algorithms. It is used widely in many research groups, companies and governments. Additionally, OpenCV is a cross-platform since it is written in C++ language which supports many platforms (Windows, Linux etc) and works with different types of compilers such as C/C++, JAVA, MATLAB and Python [149].

We create a graphical user interface (GUI) for analysing real-time applications and videos, where the face is automatically detected using the Viola-

Jones algorithm [24] and then applying a dense algorithm [58] with our ROI. Our GUI is divided into three main parts:

I. Optical flow map (Figure 4-5)
II. AU detection array for image sequences (Figure 4-6)
III. Emotions detection window for real-time application (Figure 4-7)

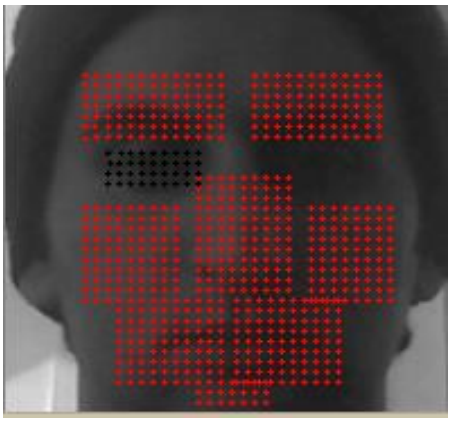


Figure 4-5: Optical flow map.

Figure 4-5 shows the GUI of the detected face using the Viola-Jones algorithm with 18 different ROI applied.

Figure 4-6: Action unit window.

Figure 4-6 shows a part of the action unit activation window; this array presents different facial features and the movement as either up, down, left or right. This array was used to identify different movements in our experiment.

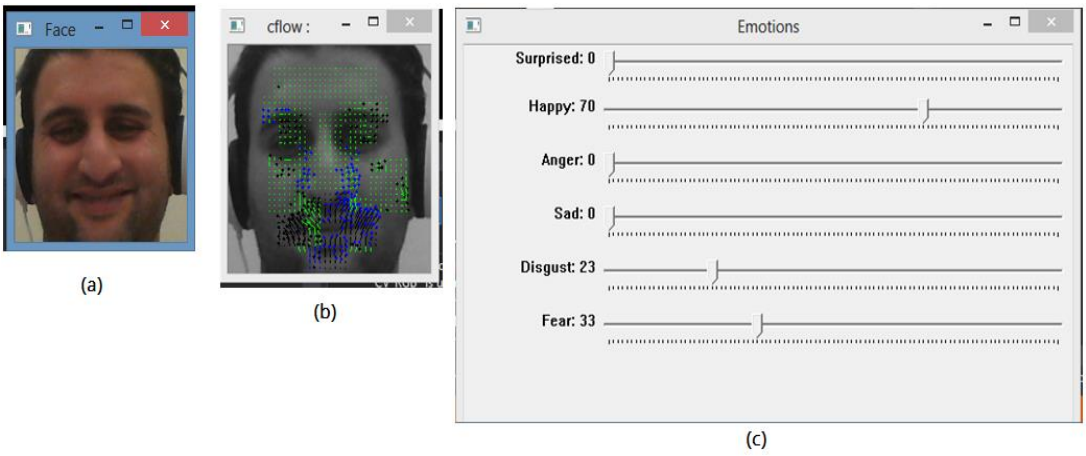


Figure 4-7: Real-time GUI.

Figure 4-7 represents the GUI for the real-time applications in our approach. Figure 4-7 (a) is the face extracted using the Viola-Jones algorithm and Figure 4-7(b) represents the ROI and optical flow map applied over the resized face. Figure 4-7(c) represents the emotions window which is a tracker bar that gives us real-time measurements of each emotion class. In this example, facial expressions are classified as follows: 70% happiness, 23% disgust and 33% fear. There are a lot of affecting factors that can justify the variation of the emotions classified in this example. First, due to the complexity of emotions and there is no pure emotion where humans can show some emotions but their subconscious reveals another emotion. Secondly, the number of emotions we have is not limited to the six basic universal emotions identified by Ekman, where the combination of multiple emotions can be expressed (as discussed in Section 1.3).

4.3 Results

We tested our approach on Cohn-Kanade Extended Facial Expression (CK+) [42] which contains 123 people expressing one or more of the seven emotions (happiness, surprise, sadness, anger, fear, disgust, contempt) where each file contains a set of image sequences that represent posed and non-posed (spontaneous) expressions from natural state until they reach the peak of the emotion. Furthermore, the CK+ database identifies additional types of metadata which include action units, facial landmarks (which have been set manually by experts), and emotion classification for each file. For more details refer to Section 1.7.

Figure 4-8 shows the detection rate of 86% for 19 action units in the CK+ dataset. By excluding AU10, AU15 and AU16 the average detection of 16 action units increases to 93%.

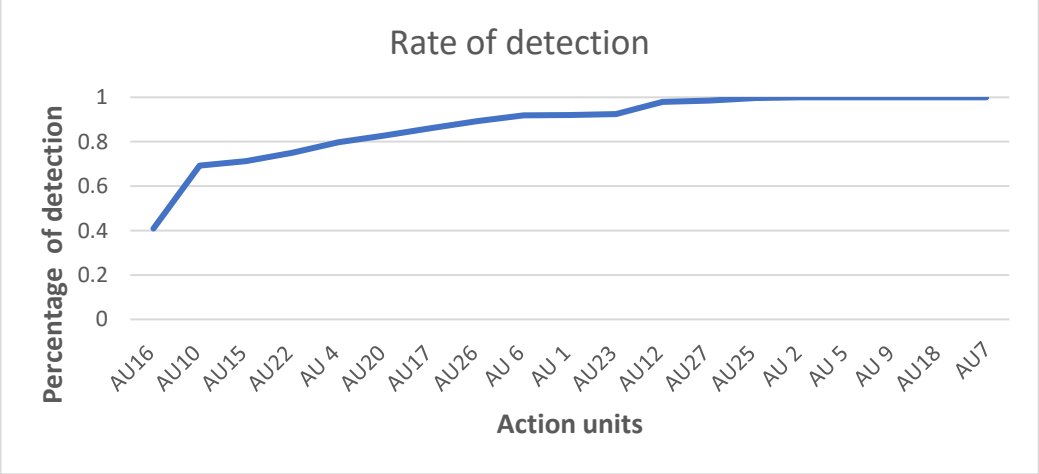


Figure 4-8: AU detection.

Figure 4-9 shows a comparison between the proposed method and a CK+ expert. In this experiment, we compute the number of occurrences for each AU denoted by the CK+ expert and compare them to the correct classification of the proposed method. Our proposed method shows an 86% agreement with the human coder.

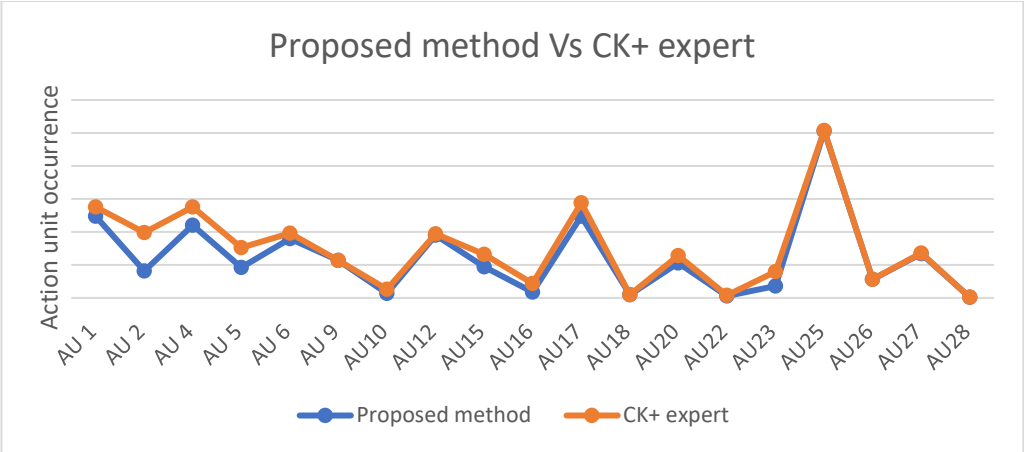


Figure 4-9: AU detection for the proposed method vs CK+ expert.

As a potential application of our work, we try to classify the six basic emotions presented by Ekman which include happiness, surprise, sadness, anger, disgust and fear. Table 4-7 shows detection percentages on the CK+ dataset.

Table 4-7: Emotions detection rate.

Emotion	Percentage
Happiness	95%
Surprise	95.2%
Anger	90%
Fear	90.5%
Sadness	82%
Disgust	70%

On average we get 87% correct classification for detecting one of the six basic emotions.

4.4 Limitations

Although our system has a high degree of accuracy and performance in detecting AU and facial expressions, testing it in a real environment produces a set of limitations which affect our detection and classification process.

4.4.1 Face Rotations and Head Movement

As we detect, face and allocated ROI face rotations will affect both the face detection, flow calculation and decision making in the MVRE. Head movements include moving forward, backward, up and down, which have a similar effect to face rotations, but since we check stabilisation of the detected face, we eliminate

some of the head movements as the stabilisation check re-initialises face detecting and re-allocating the ROI.

4.4.2 Lighting Conditions

Since our motion analysis is built using optical flow, which works on approximating pixels' intensity movement, a very bright or dark environment will affect optical flow calculation which produces failure in detecting motion.

4.5 Discussions

Recent research in action unit detection has reported a higher detection rate compared to our proposed method, such as a hybrid model (texture and geometric) which shows a 94% detection rate or using texture (Gabor function) which shows 97% or using the conventional neural network (CNN) and they gain 96%.

In this research, we present a novel algorithm which uses only facial muscles movement to classify action units. Our proposed algorithm uses a minimal set of features (352 features) to present action units compared to the state-of-the-art algorithm while maintaining a decent detection rate. Furthermore, it shows how much information about action units is encoded within the motion model. Additionally, the proposed method can be used as a pre-processing step for creating CNN features and to enhance the detection rate.

4.6 Conclusions

In this research, we present an automatic action unit and facial expression recognition system. Many other researchers have tried to solve this problem by either detecting landmarks or shapes of the facial features (eyebrows, eyes, nose and mouth). In our approach, we use the approximate location of the facial feature and try to analyse the changes/motions in these locations using the dense optical algorithm to detect activation of facial action units.

In our approach, we detect faces using the Viola-Jones algorithm [32] and normalise the face detection method using the facial normalisation parameter (FNP) in order to increase stability. Furthermore, we identify facial features location using ROI, which approximates the facial features location. ROI location is created based on studying facial muscle structure and the physiological study of analysing emotions using EMG in which we study the location of the electrode used in the EMG measurement, using dense optical flow algorithm [49] to detect facial changes caused by expressing emotions. Finally, we introduce the Motion Vector Re-Calculation Engine (MVRE) which consists of a connected web of ROI, voting system and motion converter. MVRE is used to connect the motion vector produced by optical flow to facial action units.

For classification, we use a rule-based system and motion profile, gaining a detection rate of 86% for 19 action units and 88% correct classification rate on the Cohn-Kanade Extended Facial Expression (CK+) dataset [65] that includes classification of six basic emotions (happiness, surprise, fear, sadness, anger, disgust).

Our system has shown performance ability to analyse 25 frames-per-second and detect action units in a stable environment. On the other hand, one of the challenging problems of our proposed system is face rotation, head movement and bright environmental conditions which can affect the decision-making process and the optical flow computations. These factors have been reported in a lot of recent work in analysing facial expressions discussed in Literature review Section 2.1.

Future work includes enhancing the classification method and enhancing the system, either by eliminating challenging factors and introducing a hybrid system which includes motion, texture and geometric features to gain a higher classification rate.

Finally, this approach can be used to detect different movement in each ROI related to each facial feature. We use the concept behind this approach and apply it to identifying the relationship between facial features and accurately identifying genuine and posed smiles which are the subject of discussion in the following chapter.

5 A Genuine Smile is really in the Eye – The Computer Aided Non-Invasive Analysis of Human Smiles

Several recent studies indicate that the smile represents a powerful signal for affiliative behaviour, social bond, health and longevity. Additionally, measuring the smile can predict fight outcome[150], divorce[151] and gender[90, 101] . On the other hand, a smile can be considered either as a true sign of enjoyment (real) or as negative emotions (fake, non-enjoyment). A fake smile (negative emotions) had been used in social obligated situation or as a sign of politeness, shyness, or embracement. According to [152], there are 18 different types of smiles. Each smile corresponds to a specific situation and reflects a different type of emotion.

There are two main types of the smile: real (Duchenne) and fake (non-Duchenne) [74]. According to [73], we can recognise Duchenne by raised mouth corners and raised cheeks with the appearance of eye wrinkles, whereas the non-Duchenne smile only has raised mouth corners. In terms of the FACS system, the Duchenne smile has action units (6+12) and the non-Duchenne smile has action units (12).

A lot of research supports the hypotheses of the Duchenne marker which can be used to distinguish between a real (Duchenne) and fake (non-Duchenne) smile[153-155]. Furthermore, it has been used as an evaluation tool for personal attributes such as personality, humour, loveliness and more positive [83]. Similarly, smiles inclosing this marker are regularly characterised as “true enjoyment”, “happier”, “real or felt”, “really happy” or “more genuine”[85, 156, 157].

Inspired by the research on smile analysis from both the physiological and computing perspectives mentioned in Section 2.2, we noticed some unanswered questions. Firstly, is there a weight for each facial feature in smile formulation? Secondly, which facial feature contains more information about the fake or real smile? In this research, we investigate the weight of each facial feature (eyes, cheeks and mouth) based on the changes in their location during a smile. Our approach was tested to see if there is different facial feature weight distribution in both Duchenne and non-Duchenne smiles.

The application of detecting fake and real smile can be used as a lie detecting technique in different situations such as interview and customer feedback. Additionally, our finding can be used as rules to train both human and machine to distinguish between the two types of smiles. Finally, it helps the machine to become more intelligence by looking for certain clues regarding getting instant feedback from the human by looking and their expression in different areas such as: completing tasks, evaluating a product, health and more.

5.1 Methodology

To detect the changes in facial features, we designed an automated system that measured the movement occurrences and displacement in specific regions of the face. Figure 5-1 shows the system flow which contains three main sections: detection, analysis and output. The detection phase includes: face detection, face re-size, landmark detection and identifying the region of interest (ROI). Face detection is done by using the Viola-Jones algorithm[24]. After detecting the face, we re-size it to 448 by 448 in order to apply unified ROI sizes and landmark detection for all subjects in both datasets. Landmark detection is

done using the CHEHRA model [158], which is a machine-learning algorithm used to detect 51 facial landmarks as shown in Figure 5-2 (a). Landmark detection was used to define the facial feature (eyes, cheeks and mouth) location and helped to identify ROI.

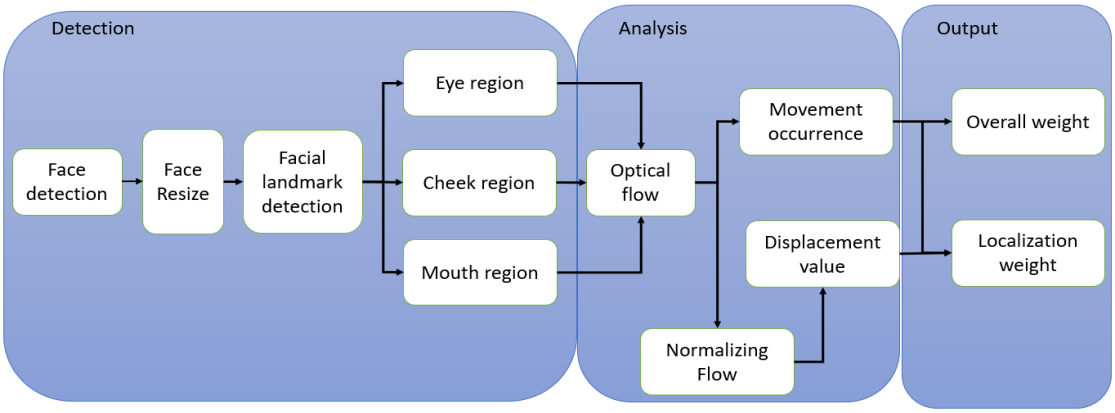


Figure 5-1 : Proposed framework.

5.1.1 Region of Interest

Identifying ROI is done by locating the facial landmarks in the neutral expression, which identify the initial location of the mouth, cheeks and eyes. This helps us to track the change and displacement value in each ROI through the smile expressions. Table 5-1 shows the ROI specification which has been applied to both datasets. Allocating ROI is done by applying two steps. First, locating two reference points where each point is related to one axis. Second, using (λ_x, λ_y) which denotes the shift-distance for each axis from the reference points. The use of (λ_x, λ_y) is to compute the ROI origin coordination point based on the reference point. Equation (5.1) shows computing of the origin coordination of each ROI origin point related to the re-sized face. After locating the origin point a pre-

defined window will be allocated with specific width and height as shown in Table 5-1.

$$ROI - Org_{(x,y)} = \begin{cases} ROI_x = P_x \pm \lambda_x \\ ROI_y = P_y \pm \lambda_y \end{cases} \quad (5.1)$$

where $\boldsymbol{ROI - Org_{(x,y)}}$represents ROI origin point coordination. $\boldsymbol{P_x}$ shows the reference point x-axis value and $\boldsymbol{P_y}$ shows the reference point y-axis value. Both values will be either added or subtracted to the $\boldsymbol{\lambda_x}$ based on ROI location in the face (left (-), right (+)). For an origin point for an edge-less area like cheeks (R9, R10, R11, R12) and below the eyes (R7, R8) area, where landmark cannot be identified, we use the nearest landmark as a reference. As an example of locating left eyebrows, using landmark $\boldsymbol{P_1}$ (Figure 5-2 (a)), we allocate a window of 110 by 35 which covers the area of the eyebrows. The size of each ROI has been set to cover all possible occurrences and displacement movement. As an example of locating right cheek R12, we use mouth right corner point (Figure 5-2(a), $\boldsymbol{P_{38}}$) as a reference point to compute cheeks origin point, as well as R10, R9 and R11.

The specification of each ROI shown in Table 5-1 is set after resizing the face to (448*448 pixels) then trying different ROI window size. The determination of the window size done by conducting a total of 360 experiments was carried out to identify each ROI size. The size of each ROI is taken by overfitting window to cover all possible movement of the corresponded facial feature which fits both the CK+ and the MUG dataset. This done to unify the analysis between the two datasets and apply the motion algorithm within the same size location for each facial feature for both datasets.

Table 5-1 : ROI specifications.

Facial features	ROI	Size (width, height) in pixels	Reference landmark	λ_x	λ_y
Eyes	R1	110,35	$P_{1,X}, P_{2,Y}$	20	20
	R2	110,35	$P_{6,X}, P_{7,Y}$	20	20
	R3	45,50	$P_{12,X}, P_{1,Y}$	20	20
	R4	30,60	$P_{11,X}, P_{12,Y}$	15	15
	R5	30,60	$P_{17,X}, P_{18.Y}$	15	15
	R6	45,50	$P_{20,X}, P_{19,Y}$	20	20
Cheeks	R7	25,50	$P_{11,X}, P_{16,Y}$	15	10
	R8	25,50	$P_{17,X}, P_{22,Y}$	15	10
	R9	60,60	$P_{32,X}, P_{32,Y}$	60	90
	R10	60,60	$P_{38,X}, P_{38,Y}$	60	90
	R11	60,60	$P_{32,X}, P_{32,Y}$	60	30
	R12	60,60	$P_{38,X}, P_{38,Y}$	60	30
Mouth	R13-R18	20,20	$P_{32(X,Y)}, P_{33(X,Y)}, P_{35(X,Y)}, P_{37(X,Y)}, P_{38(X,Y)}, P_{39(X,Y)}, P_{41(X,Y)}, P_{42(X,Y)}$	20	20

Using facial landmarks as a reference point to allocate each ROI (Figure 5-2 (b)) we identify 18 regions of interest which cover the area of each facial feature. The eye region is covered by four different areas: eyebrows (R1, R2), eye (R4, R5), eye corner (R3, R6) and the area underneath the eye (R7, R8). As has been proved by many research [73, 80, 159], the difference between a real and fake smile is the eye region. Therefore, to test the hypothesis of distinguishing between a fake and real smile we divided the eye region into different ROI.

Cheeks areas were divided into two main segments left cheek (R9, R11) and right cheek (R11, R12) and mouth area was divided into six main segments upper lip left (R13), upper lip middle (R14), upper lip right (R15), lower lip left (R16), lower lip middle (R17) and lower lip right (R18). The segmentation helps us analyse each part of the facial features movement through smile expression and how it contributes in smile formulation.

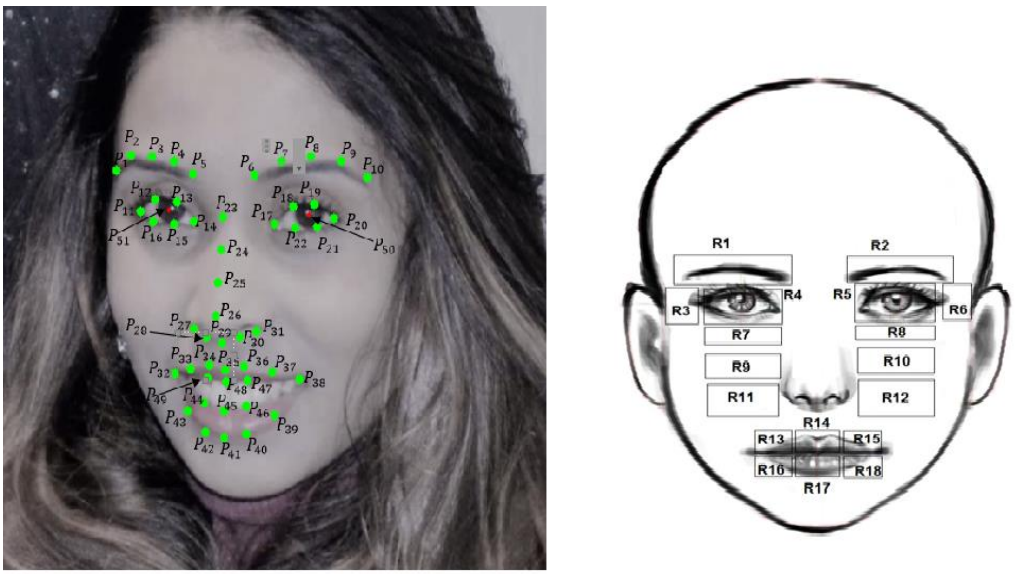


Figure 5-2 : (a) Landmarks detection using the CHEHRA model, (b) Region of interest.

5.1.2 Computing Movement using Optical Flow

Since we try to imitate the EMG way of analysing the muscle movement using computer vision techniques. We try different motion analysis techniques which include background subtraction, image subtraction and optical flow. Background subtraction does not work with the face as it needs a background to measure the motion of an object, as the face and the facial feature does not have background model this model failed to detect the motion[160]. Image subtraction algorithm can be used to analysis facial movement but it very noisy and sensitive to light [161].

Finally, we use optical flow represent by applying Gunnar Farnebäck's [147]. The optical flow algorithm is a two-frame motion estimation algorithm, where Farnebäck uses quadratic polynomials to approximate the motion between two sequences frames to approximate each neighbourhood pixels movement.

Comparing optical flow to the other algorithms we found that it represents the motion better and has lower sensitivity to light and noise. Furthermore, there is two version of the optical flow algorithm: Farnebäck's and Lucas and Kanade's algorithm [58]. Farnebäck's algorithm is considered to be a dense optical flow algorithm since it computes the flow for all pixels in the image which retrieve more accurate motion data than a sparse model present by Kanade's algorithm, for more details refer to Appendix B. Using Farnebäck's algorithm, we can estimate the displacement value and movement occurrence for each facial feature.

After applying optical flow, we normalise the displacement value to overcome some challenging factors like face location to the camera to the face and face size. Furthermore, to unify comparison between the two datasets based on the displacement value, we normalised the displacement value in each ROI by dividing it with the triangle constructed from the tip of the nose (Figure 5-2 (b), $\boldsymbol{P_{29}}$) and the eyes corners (Figure 5-2 (b), $\boldsymbol{P_{11}, P_{20}}$). After normalising the displacement value, we eliminate some fault movement caused by optical flow computation, which will affect measuring the movement occurrence or total displacement. To eliminate this fault movement, we check the displacement value of each region of interest $\boldsymbol{f_{ROI}}$ and compare it to $\boldsymbol{\varepsilon}$ which is set to (0.1) as shown in equation 5.2. $\boldsymbol{Mo_{ROI}}$ represents the movement occurrence as the Boolean value set to (true) if the flow value exceeds the $\boldsymbol{\varepsilon}$ which is used later in calculating the weight distribution based on the movement occurrence,

$$\boldsymbol{if\ f_{ROI} > \varepsilon :\ Mo_{ROI} = True,} \tag{5.2}$$

The value of ε was set by carrying out a set of experiments done on checking each ROI related to the facial feature (mouth, nose, eyes) in a video

with a non-moving face. This done by creating 10 videos (5 from CK+ and 5 from MUG dataset) contains the first frame repeated for a 5 second. These videos were created to test the optical flow calculation on a non-moving face. By trying different ε value we set to 0.1 for both datasets. This false movement is caused by the optical flow approximation equation in computing the movement.

5.1.3 Smile Weight Distributions

Subsequently, as shown in Figure 5-1, the last part of our proposed work is output, where we study the relation between facial features in fake and real smiles in two levels: overall weight and localisation weight. Overall weight represents the relation between the main facial features and the fake and real smile. Localisation weight represents the relation between different parts of the facial features and fake and real smiles. Overall weight is computed based on movement occurrences and displacement value while localisation weight computed is based on displacement value.

Computing overall weight is based on the occurrence of movement done by using equation 5.2, where we checked whether the value in the region of interest area is higher than $\boldsymbol{\varepsilon}$, which indicates movement and is considered as the score for the corresponding facial feature. The weight is computed for each facial feature by counting the number of movement occurrences to the number of subjects which indicates how many subjects used the specific facial features through smile expression.

Computing overall weight is based on the total displacement done by applying equations 5.3 to 5.10. Equation 5.3 represents computing the flow in specific ROI, where $\boldsymbol{Td_i}$ represents total displacement in ROI with ID $\boldsymbol{i}$. This is

done by checking the flow value using equation 5.2 then summing up all displacement values of the flow in each possible direction which represent the total displacement in specific ROI as shown in equation 5.3. We sum up all the displacement values of the flow in each possible direction since our analysis focuses on analysing the smile from natural to the peak of the smile.

$$Td_{R(i)} = \sum f_{(R(i),Up)} + \sum f_{(R(i),Down)} + \sum f_{(R(i),Left)} + \sum f_{(R(i),Right)} . \quad (5.3)$$

Equations 5.4, 5.5 and 5.6 represent the total displacement computed using ROI for both eyes $\boldsymbol{Ef}$, cheeks $\boldsymbol{Cf}$ and mouth area $\boldsymbol{Mf}$ respectively. Equation 5.7 represents the total displacement of the face $\boldsymbol{Td_{Face}}$; this is done by summing up the displacement in the eyes, cheeks and mouth.

$$Ef = \sum_{i=1}^{6} Td_{R(i)} . \quad (5.4)$$

$$Cf = \sum_{i=8}^{12} Td_{R(i)} . \quad (5.5)$$

$$Mf = \sum_{i=13}^{18} Td_{R(i)} . \quad (5.6)$$

$$Td_{Face} = Ef + Cf + Mf . \quad (5.7)$$

After computing equations 5.4 – 5.7 we gain the flow value normalised for each facial feature. Overall weight is computed by dividing the displacement value of each facial feature by the total displacement as shown in equations 5.8, 5.9 and 5.10 for eyes, cheeks and mouth respectively.

$$Eye_{weight} = {}^{Ef}/_{Td_{Face}}. \quad (5.8)$$

$$Cheek_{weight} = {}^{Cf}/_{Td_{Face}}. \quad (5.9)$$

$$Mouth_{weight} = {}^{Mf}/_{Td_{Face}}. \quad (5.10)$$

Computing localisation weight ($\boldsymbol{Loc - weight_{R(i)}}$) for each ROI is done by dividing ROI displacement value $\boldsymbol{Td_{R(i)}}$ by the corresponding facial feature total displacement value $\boldsymbol{Td_{ff}}$. This indicates how much each ROI contributes to the movement of the facial feature through smile expression. Localisation weight distribution ($\boldsymbol{Loc - weight_{R(i)}}$) can be computed using the following equation,

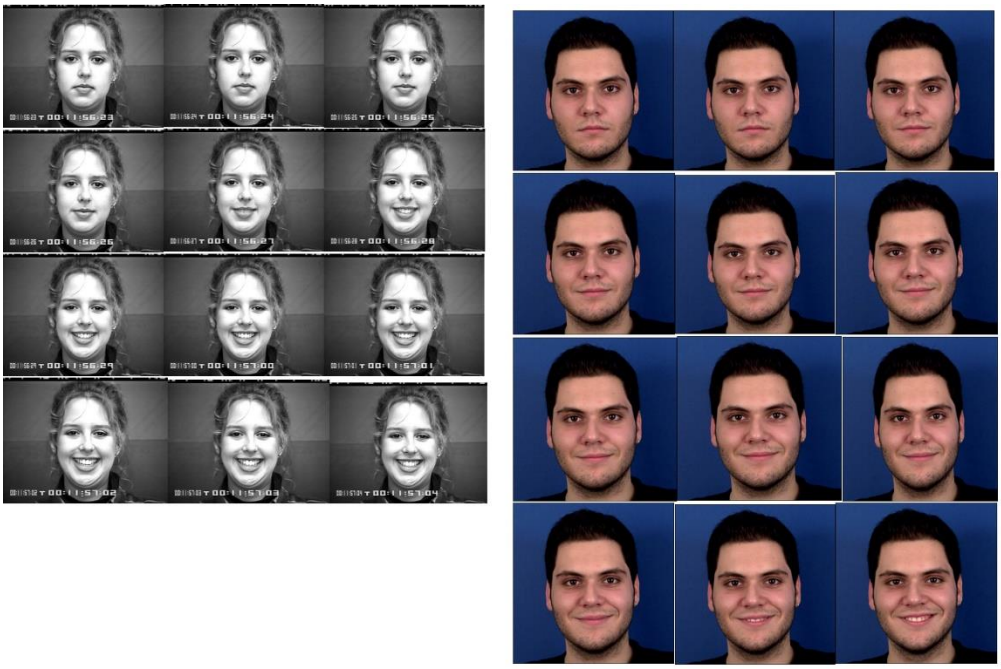
$$Loc - weight_{R(i)} = {}^{Td_{R(i)}}\Big/_{Td_{ff}} \quad where\ R(i) \in ff. \quad (5.11)$$

5.2 Experiment Setup

In this section, we describe our experimental setup. We use two different datasets where each represents subjects expressing fake or real smile. The CK+ dataset represents a 97-subject expressing posed emotions, where each subject is set facing the camera with fixed lighting. The subjects will be asked to express emotions. The CK+ dataset contains metadata which identify: actions units, landmarks and six emotions (happy, surprise, anger, fear, disgust and sad). A total of 486 images were used to analysis the fake smile.

The MUG dataset contains 52 subjects of Caucasian origin aged between 20-35 expressing laboratory-induced emotions. In order to create non-posed emotions subjects were recorded while they were watching a video to help them to induce emotions. For video recording setup, subjects were sitting in a chair in front of the camera with a blue background with two 300w light sources. The camera has recorded a video with 19 frames per second with size

896×896 pixels. Figure 5-3 shows a sample of the image sequence used in our experiment, Figure 5-3 (a) represent subject from the CK+ dataset and Figure 5-3(b) shows from the MUG dataset.



(a) (b)

Figure 5-3 : Image sequences (a) CK+ dataset and (b) MUG dataset.

5.3 Results

As stated before we use optical flow to measure two different attributes: movement occurrence and displacement value.

5.3.1 Displacement Values

In this experiment, we measure flow value for the first three frames and the last three frames of the smile for both datasets. The first three frames are denoted by NF_1, NF_2, NF_3 which represent the neutral expressions or the start of the smile. The last three frames are denoted by PF_1, PF_2, PF_3 which represent frames in the peak of the smile, using dense optical flow algorithm to measure flow which represents the displacement value of pixels in the related region of interest (ROI) for each of the facial features.

Figures 5-3, 5-4 and 5-5 show two random subjects from both datasets and the corresponding facial features and flow value. Figure 5-3, Figure 5-4 shows the mouth and cheeks flow respectively. Both figures show that MUG dataset has a slightly higher flow value compared to CK+ dataset. On the other hand, Figure 5-5 represents eyes flow and shows the big difference in flow value between MUG and CK+ datasets where the subject in the MUG dataset shows higher flow value.

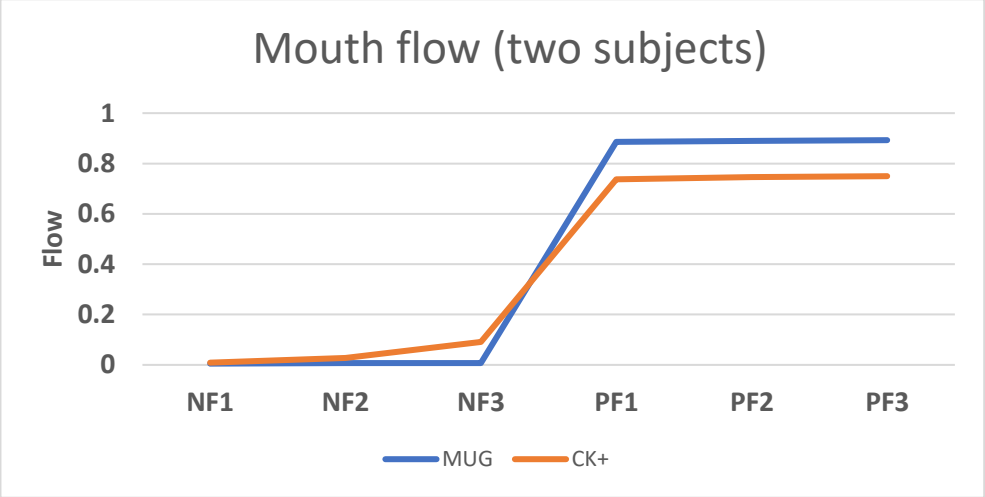


Figure 5-4: Flow around mouth for two subjects.

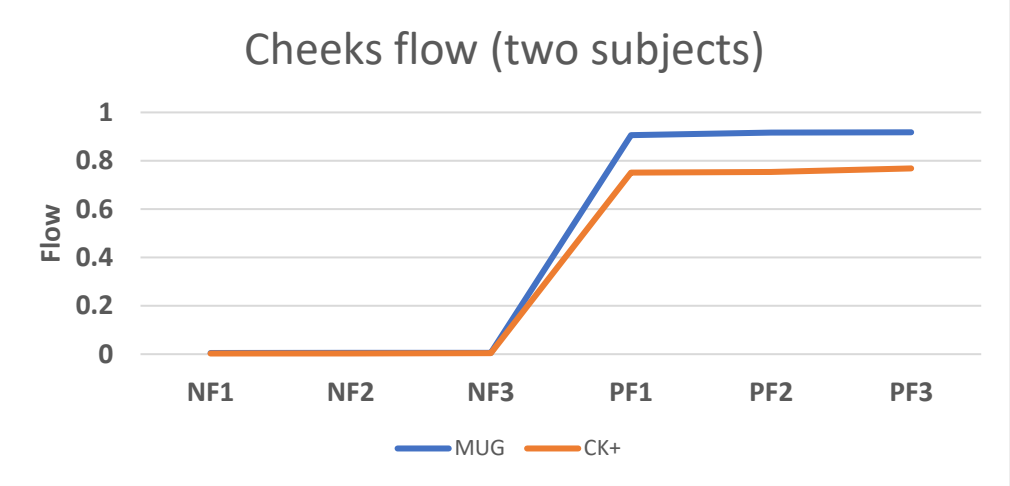


Figure 5-5: Flow around cheeks flow for two subjects.

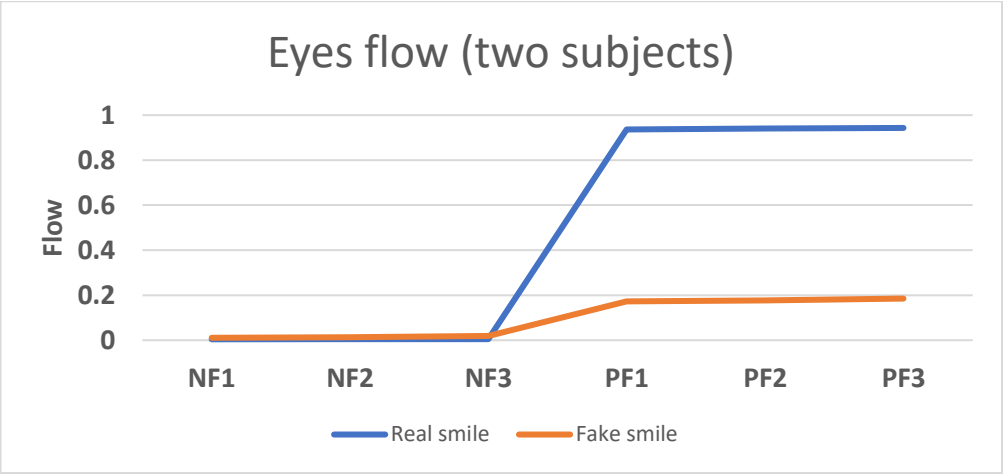


Figure 5-6: Flow around eyes for two subjects.

In order to check if these values have a significant meaning, we compute the average of each frame in neutral and peak frames (NF_{1-3}, PF_{1-3}) for all the subjects in the datasets using the following equations,

$$Average\_NF_i = {}^{1}/_{N} \sum_{j=1}^{N} NF_{i,j}, \quad (5.12)$$

$$Average\_PF_i = {}^{1}/_{N} \sum_{j=1}^{N} PF_{i,j} \ , \quad (5.13)$$

where $NF_{i,j}$ represents the neutral flow value of frame i at subject j. $PF_{i,j}$ represents the peak flow value of frame i at subject j and N denotes the number of subjects.

Figure 5-6, 5-7 shows the average of the neutral and peak frames for mouth and cheeks. Both figures show that MUG datasets have a slightly higher average compared to the CK+ dataset. Figure 5-8 shows the eye average flow for both datasets which shows a significant difference between MUG and CK+

datasets.

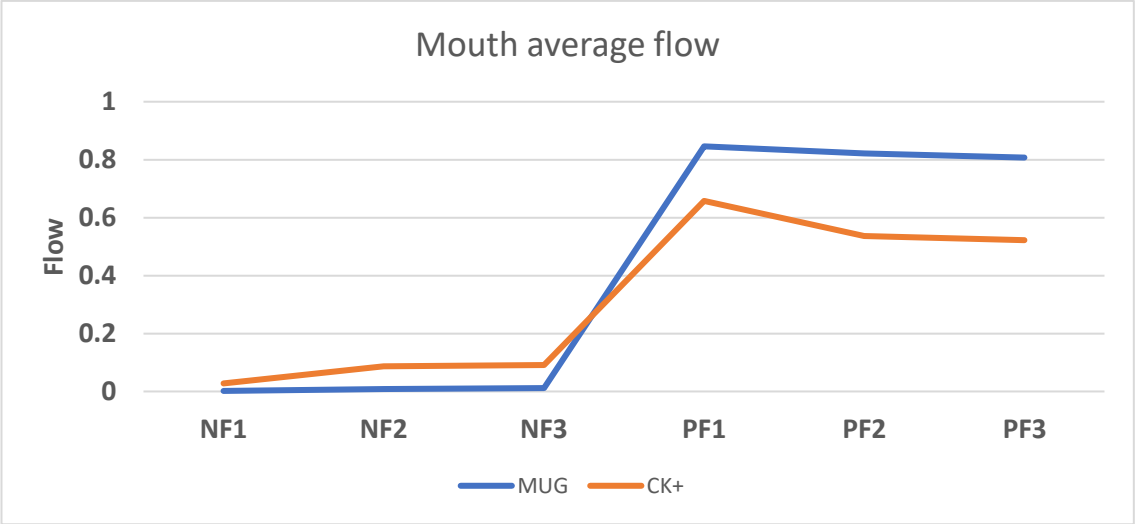


Figure 5-7: Flow around the mouth flow for neutral and peak frames.

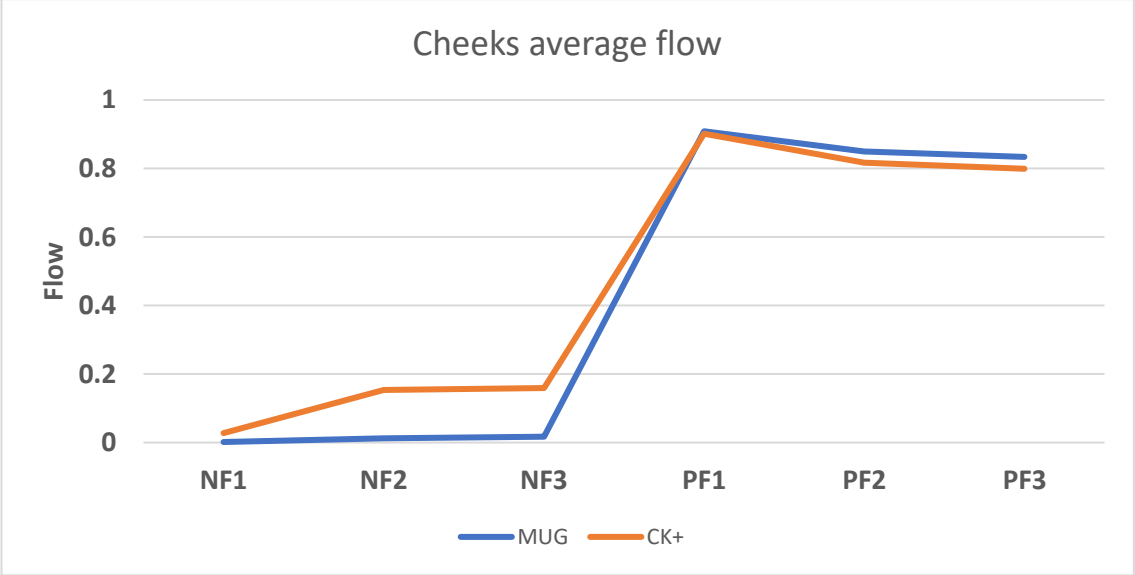


Figure 5-8: Flow around the cheek for neutral and peak frames.

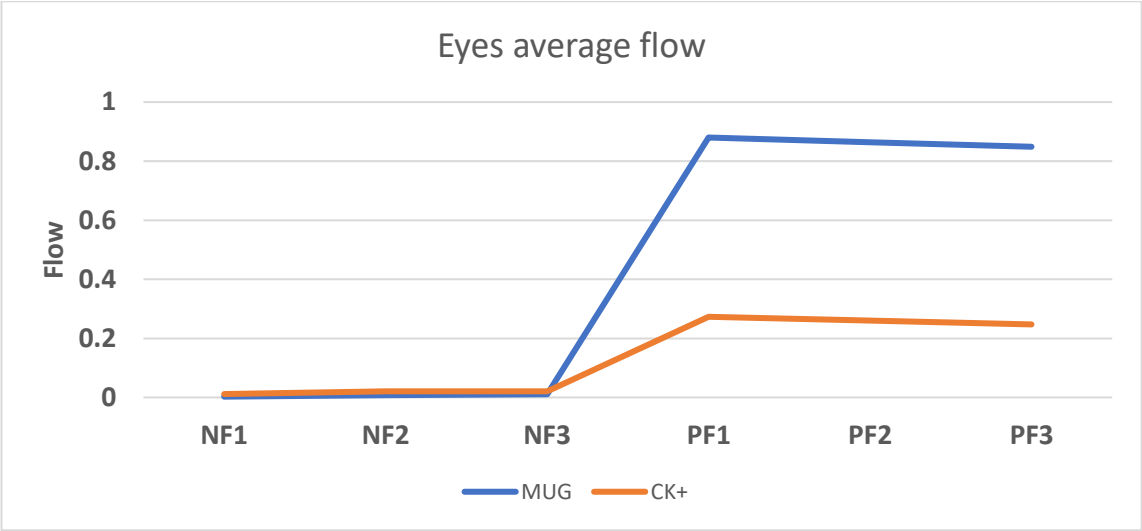


Figure 5-9: Average flow around the eyes for neutral and peak frames.

Furthermore, we compute the median of each frame in neutral and peak frames (NF_{1-3}, PF_{1-3}) for all the subjects in the datasets. Figures 5-9, 5-10 show the median of the neutral and peak frames for mouth and cheeks. Both figures show that MUG datasets have a higher median value compared to the CK+ dataset. Figure 5-11 shows the eye median flow for both datasets which shows a significant difference between MUG and CK+ datasets which approve with the result computed by the average.

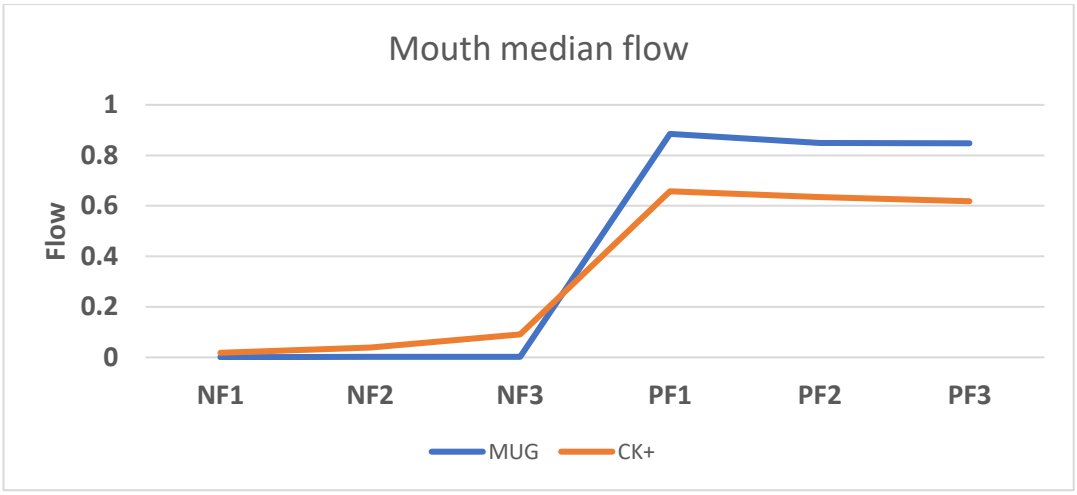


Figure 5-10: Median flow around the mouth for neutral and peak frames.

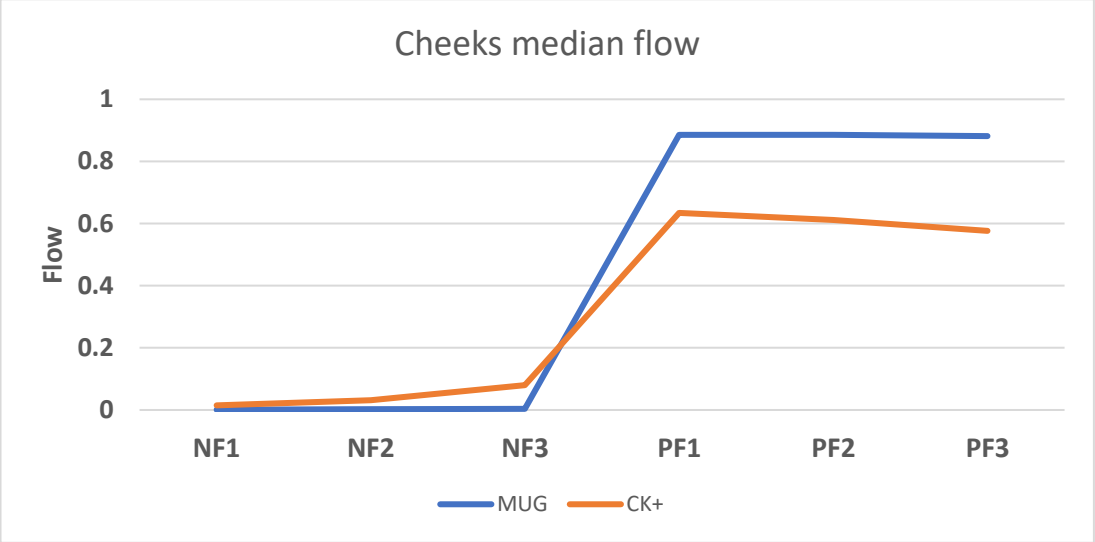


Figure 5-11: Median flow around the cheeks for neutral and peak frames.

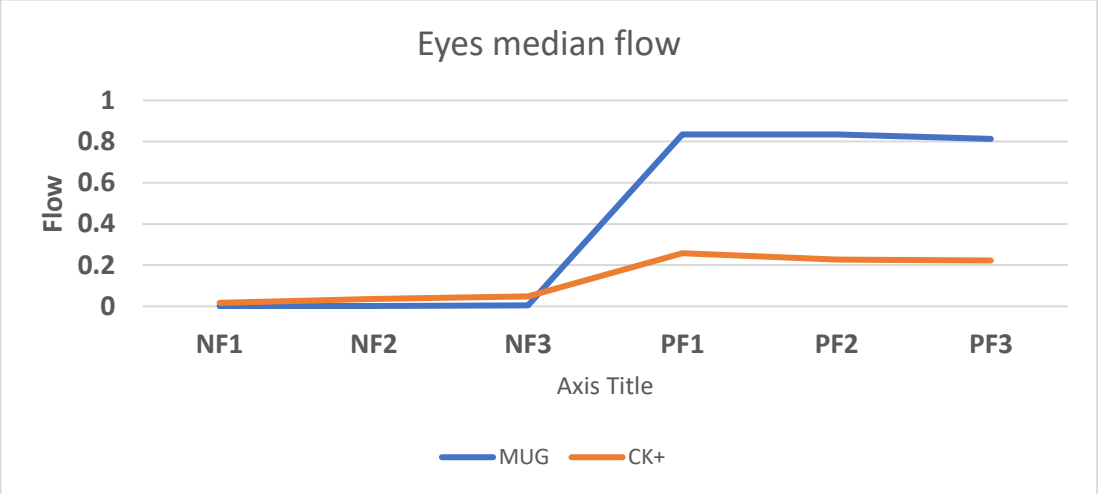


Figure 5-12: Median flow around the eyes for neutral and peak frames.

In the second experiment, we compute the peak frames average (PFA) for all subjects in both datasets. In order to show the average difference between these datasets in the peak of the smile, PFA is computed using the following equation,

$$\text{PFA} = {}^{1}/_{(3 * N)} \sum_{j=1}^{N} PF_{1,j} + PF_{2,j} + PF_{3,j}, \qquad (5.14)$$

where $PF_{1,j}, PF_{2,j}, PF_{3,j}$ represent the first, second and third peak frames of subject j respectively. N represents the number of subjects in the datasets; we divide by $(3 * N)$ since we sum up the three-peak frame for each subject.

Figures 5-12, 5-13 and 5-14 show the PFA and standard deviations for all the subjects in both datasets for mouth, cheeks and eyes respectively. As shown in these figures, there are differences between the MUG and CK+ dataset in mouth and cheeks where mouth in the MUG dataset has an increase of 164% of flow value and cheeks shows 169% increase of flow compared to the flow value in CK+ dataset. Finally, the eyes in the MUG dataset show an increase of 412%. Figure 5-15 summarises the increased percentages of each facial feature in the MUG dataset compared to the CK+ dataset.

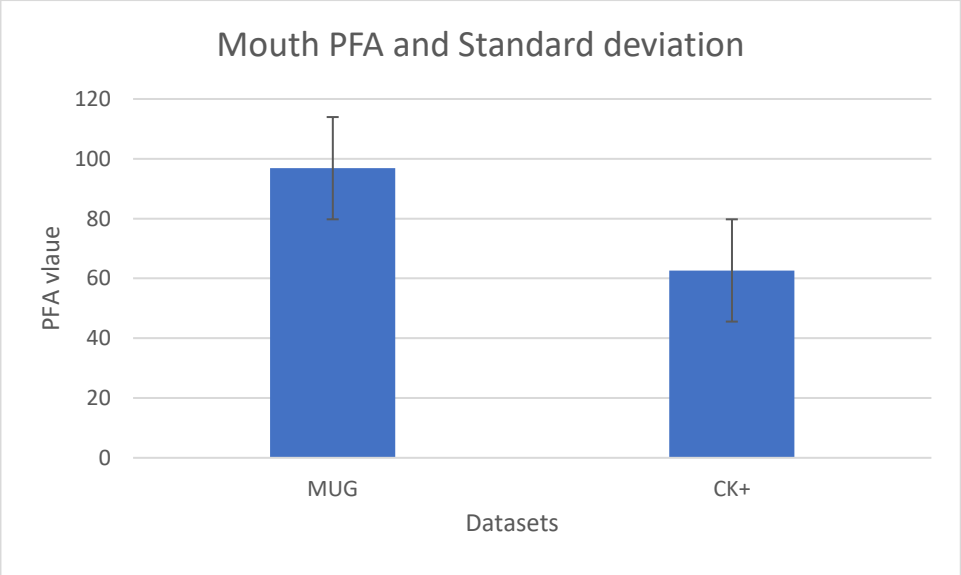


Figure 5-13: Mouth PFA and SD in MUG and CK+ datasets.

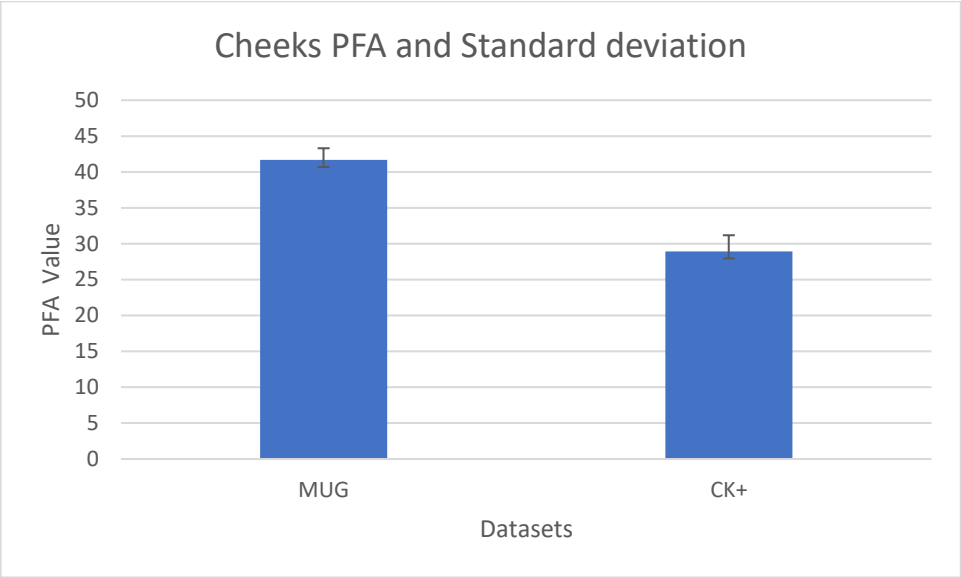


Figure 5-14: Cheeks PFA and SD in MUG and CK+ datasets.

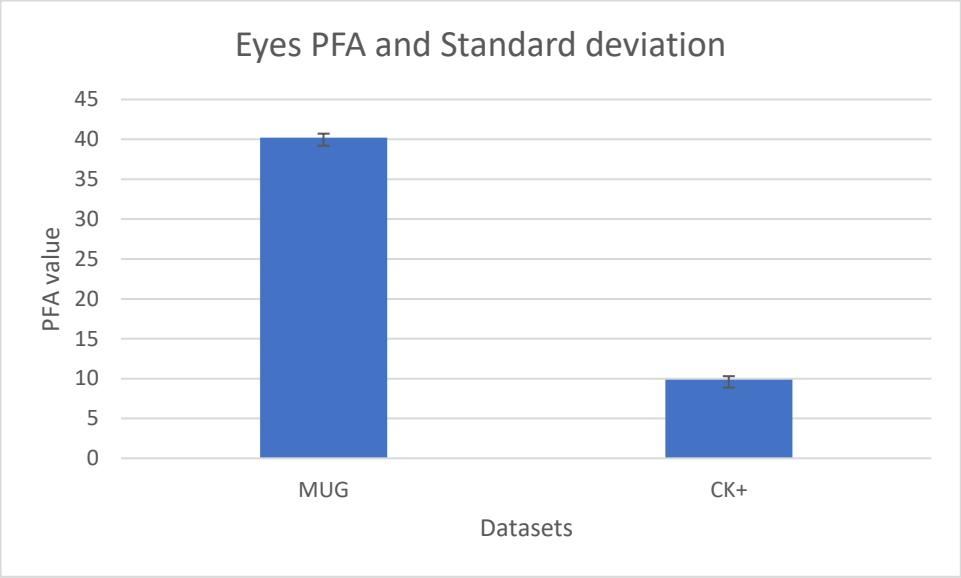


Figure 5-15: Eyes PFA and SD in MUG and CK+ datasets.

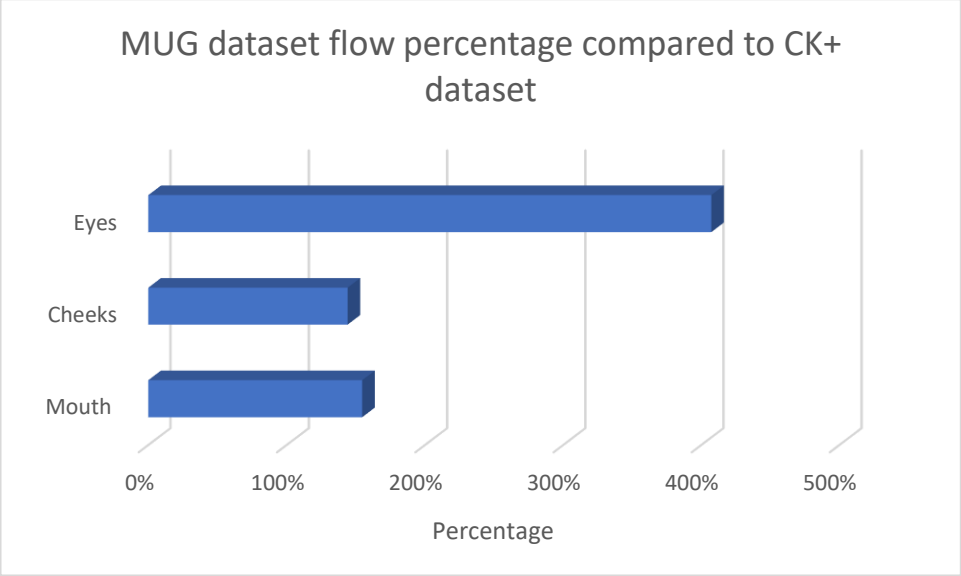


Figure 5-16: MUG dataset flow percentage for each facial feature compared to CK+ dataset.

Finally, Figure 5-16 shows the distributions of flow for the region of interest in both eyes in peak frame for both datasets. As shown, the MUG dataset has a higher mean and median compared to CK+ dataset. Furthermore, it shows MUG dataset has more distributed flow value compared to CK+ dataset.

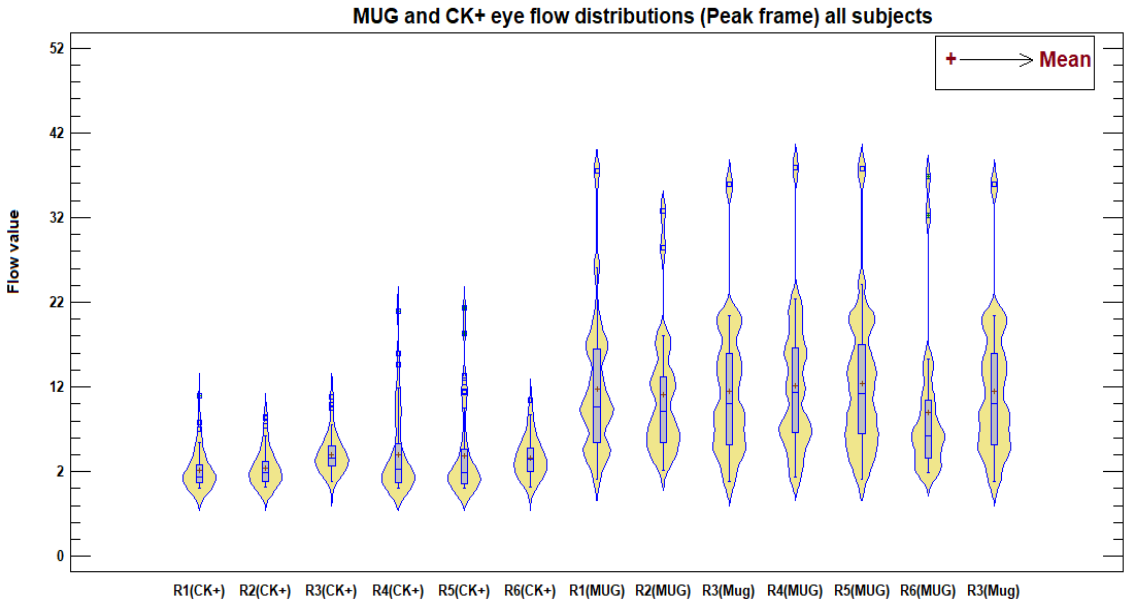


Figure 5-17: Eyes' ROI flow distributions for both datasets.

To test the hypotheses that fake and real smile is associated with statistically significant mean an independent samples T-test was performed on the neutral interval (NF1, NF2, NF3) and peak interval (PF1, PF2, PF3) for each facial feature. First, a Levene test was performed on the data in order to test the homogeneity of the variances to the null hypotheses for each facial feature and time interval. As shown in Table 5-2 represent different P-value which obtained from the Levene test which determines the equal or unequal variances to be used in the T-test.

Table 5-2 : Levene-test on facial features.

Facial features	**P-value at neutral interval**	**P-value at peak interval**
Mouth	0.003413	7.39E-05
Cheek	8.59E-30	0.160738
Eye	1.67E-13	0.000313

Applying the T-test shows that there is a significate difference in the score of the eyes flow (peak interval) in a real smile ($M = 40.2,\ SD = 0.51$) and a fake smile ($M = 9.87\ ,\ SD = 0.42$) with the conditions ($T(73) = -2.55\ , p_{\text{Peak}} = 0.0128$) assuming unequal variances based on Levene test. Other facial features showed no significate difference in both time intervals (neutral, peak). The results of the T-test carried on the mouth and the cheeks as follows:

The mouth area (assuming unequal variances):

$T(73) = -4.005$, $p_{\text{Peak}} = 0.06147$,

$T(73) = 3.89$, $p_{\text{neutral}} = 0.0527$.

The cheeks area (peak interval: assuming equal variances, neutral interval: assuming unequal variances):

$T(73) = -3.08$, $p_{\text{Peak}} = 0.078$,

$T(73) = 4.439$, $p_{\text{neutral}} = 0.11231$.

5.3.2 Movement Occurrences

For the second experiment, we measured the movement occurrence in both datasets. Any motion in the 18 different regions of interest was computed. Figure 5-20 summarises the percentage occurrence of the 18 different regions of interest with their corresponding facial features. The CK+ dataset showed all subjects (100%) having motion in the mouth and cheeks area, with 30% of the subjects showing movement in the eyes region. Conversely, although the MUG dataset had the same percentage occurrence in the mouth and cheeks, the eyes region showed 62%, which shows an increase of 32% in movement occurrence.

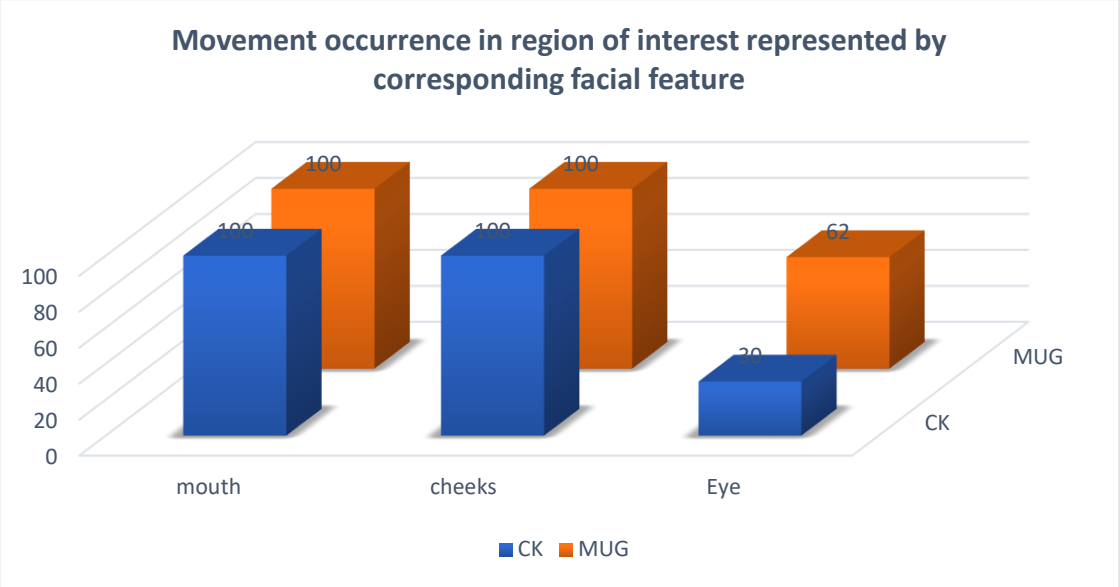


Figure 5-18 : Movement occurrence in facial features.

Figure 5-21 shows the overall total flow computed using Gunnar Farnebäck's algorithm in the 18 different regions of interest related to each facial feature; we computed the total flow using equation 5-2. The CK+ dataset showed that 73% of the total motion vector occurred in the mouth area, 21% in cheeks and 6% in the eyes area. In the MUG dataset, 66% of motion vectors occurred in the mouth area, 20% in cheeks and 14% occurred in the eyes area.

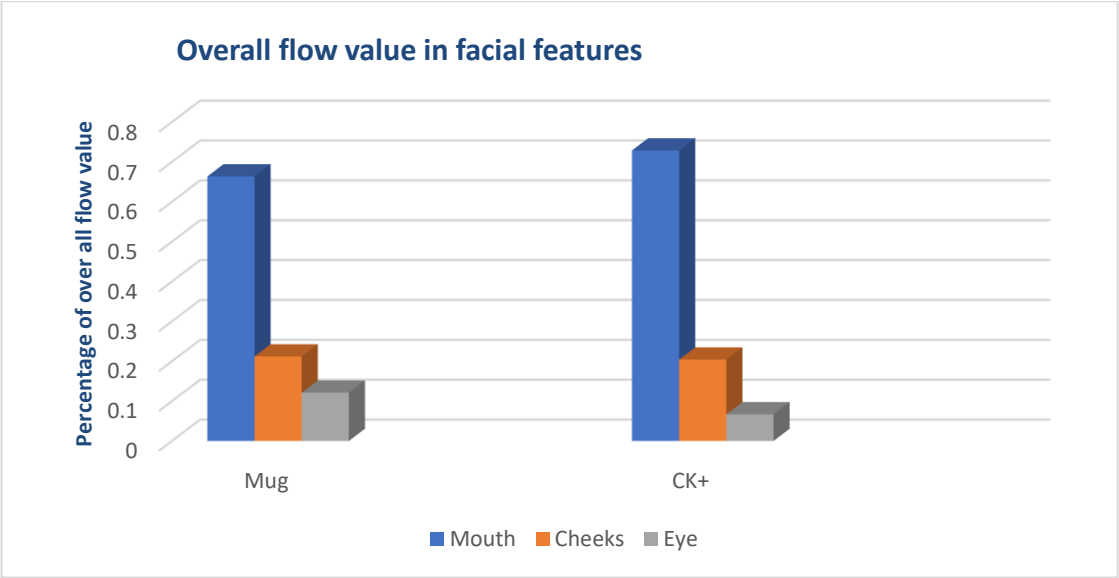


Figure 5-19: Overall flow value in facial features.

Figure 5-22 shows detailed weight distribution of the mouth, cheeks and eyes region. The CK+ dataset shows 100% of subjects with movement in the mouth and cheeks. For the eye region, as stated previously, we divided the section into three main segments. 30% of the subjects show movement in the eyes region (R7, R8), 25% showed movement in the eyebrows (R1, R2) region and 54% showed movement in the eye corner region (R3, R6). MUG dataset showed 100% of subjects with movement in the mouth and cheeks. For the eyes segment, 65% of subjects showed movement in the eyes region (R7, R8), 60% showed movement in the eyebrows (R1, R2) region and 63% showed movement in the eye corner region (R3, R6).

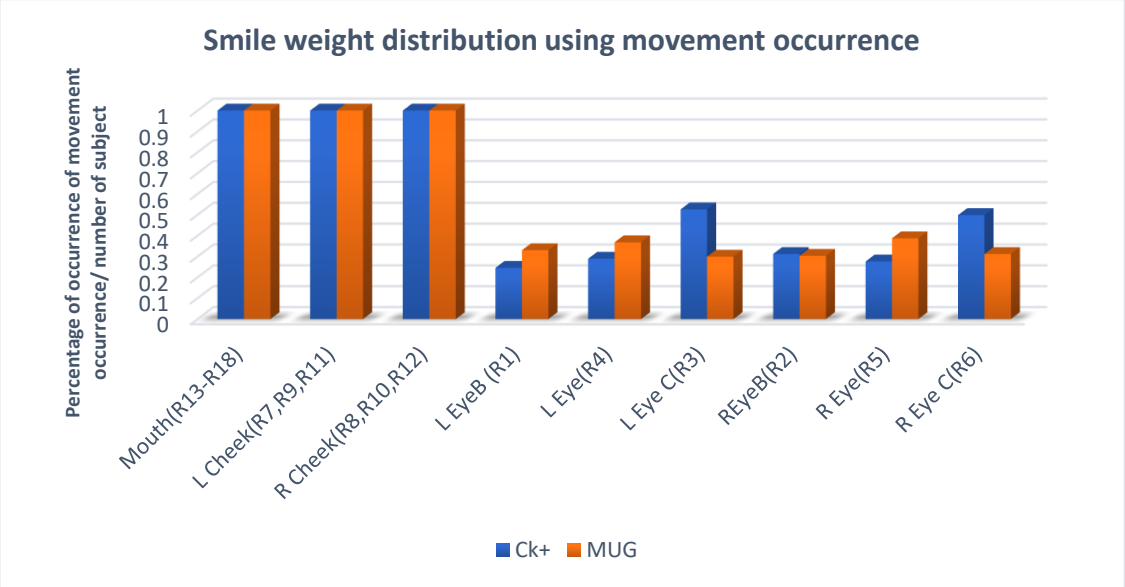


Figure 5-20 : Detailed ROI flow for facial features.

5.4 Discussions

As shown in the results, we found there is a difference in the flow values and movement occurrences in Duchenne and non-Duchenne smiles.

In the first experiment, we study the total displacement value for the first three frames (NF_1, NF_2, NF_3) and the last peak frame (PF_1, PF_2, PF_3) of the smile. The results show that both the average and the median of the facial features in a real smile are higher when compared to a fake smile. Furthermore, the results indicate that the motion around eyes have a significant difference between the real and fake smile, where it shows an increase of 4-fold in the displacement value of the genuine smile.

Additionally. We computed the motion distribution of the subjects expressing real and fake smile for each ROI related to the eyes area. The results show that subjects expressing real smile have higher distribution and median

value compared to the fake smile in each ROI of the eyes as shown in Figure 5-16.

In the second experiment, we study the movement occurrence of each facial feature in posed and genuine smiles. The results show an increase of 30% motion occurrence in subjects expressing a Duchenne smile, as shown in Figure 5-18. Furthermore, our second experiment showing the total flow value of the regions of interest, indicate that the eye region in a Duchenne smile has 13% of the total flow compared to a non-Duchenne smile which has 6% as shown in Figure 5-18.

5.5 Limitations

In this research, we analyse the relation between facial features and the genuineness of smiles. Although our result agreed with the physiological findings our research have a set of limitations. We highlight them below.

1. Head movement

Our analysis is carried out on two datasets which were recorded in controlled environments and with minimal to no head movement. Head movements in real life is very likely affect our results.

2. Dataset

The two datasets include laboratory generated emotions, which we think if we can get a dataset to containing more genuine smiles then our analysis and statistics computation will likely to be different.

3. Lighting conditions

As we use optical flow algorithm for measuring the facial muscle movement, very bright or dark lighting conditions will affect the calculation of the optical flow and it will not properly detect facial movement.

4. Creating real time application

Our method can't be used for creating an application as it very sensitive to lighting condition and head movement. On the other hand, our approach can be used as a based rule for creating a real-time application.

5. Face pose

Our analysis was performed only on faces with a frontal view and our approach cannot be applied on faces with partial or profile views.

5.6 Conclusions

In this chapter, we have looked at identifying the genuineness of a smile by means of analysing the dynamic components of the facial expression associated with a smile. As shown in the second experiment (i.e. motion occurrences), the weight distribution of the smile implies that a smile can be detected accurately in both mouth and cheeks areas since motion occurrence in these areas is 100% among the test subjects in both Duchenne and non-Duchenne smiles. Additionally, 62% of subjects expressing a Duchenne smile used their eyes when expressing a smile, whereas during a non-Duchenne smile, only 30% of subjects used the eye area, which shows we can detect smile

through eye region in a Duchenne smile with an additional 32% compared to a non-Duchenne smile.

To answer the question of what part of the facial feature contains more information with regard to a genuine smile our studies conclude that it is the eyes, as previously shown through the physiological research [12, 20 and 21]. As shown in Figure 5-18 there is a 7% increase in activity around the eye. Furthermore, a real smile shows a higher average and median value when compared to the posed especially in the eye area. Additionally, we focussed the eyes ROI where subjects expressing Duchenne smile show higher distribution in displacement value whereas in the case of non-Duchenne smiles the eyes corner have the higher weight compared to the other region. In Duchenne smiles, it shows that the distribution of the ROI in the eyes are similar which indicates that the eyes are more active when expressing Duchenne smile as shown in Figure 5-16.

Finally, taking inspiration from the results of this chapter, we study a deeper aspect of emotion where we discuss the analysis of the smile dynamic looking for clues in gender. Thus, decoding gender from the dynamic components of a smile is the topic of the next chapter.

6 Gender Identification using the Smile Dynamics

It is often said that the face is a window to the soul. Bearing a metaphor of this nature in mind, one might find it intriguing to understand, if any, how the physical, behavioural as well as emotional characteristics of a person could be decoded from the face itself. With the increasing deductive power of machine learning techniques, it is becoming plausible to address such questions through the development of appropriate computational frameworks.

Computational frameworks for human face analysis have recently found their way into great many application areas. These include computer vision, psychology, biometrics, security and even healthcare. The appealing, and the practical, nature of such face analysis techniques, are highlighted by the wealth of information it can provide in a non-invasive manner. Unsurprisingly, such applications have already found their way into furnishing useful tell-tale signs of an individual's health status, identity, beauty and behaviour, all of which can be enhanced by the non-invasive information that leaks directly from the face, e.g. [89, 92, 162].

Additionally, computer-based analysis of the human face can provide strong and useful cues for personal attributes such as age, ethnicity and more appropriately gender, in the present context. Gender classification, in this sense, can, for example, aid as an advantageous biometric feature in order to improve the accuracy of determining an identity, especially in the presence of limited information on a subject.

Recent research into gender classification has faced challenging hurdles, mainly due to the reliance on static data in the form of facial images. There are

many inherent issues when looking for gender clues in the appearance-based facial analysis. These include variability of lighting conditions, pose and occlusions. In this regard, in this work, we departed from such appearance-based analysis of facial images. Instead, we consider the analysis of the dynamic face, in particular, the dynamics of the smile, for clues of gender. This allows us to address the very intriguing question of whether a person's sexual dimorphism is encoded in the dynamics of the smile itself.

This research is concerned with the identification of gender from the dynamic behaviour of the face. Equally importantly, we seek to answer the crucial question of whether gender is encoded in the dynamics of a person's smile. The case for such a computer-based investigation is fuelled by an array of cognitive physiological studies showing evidence of gender variances in facial expressions, e.g.[93-95]. We specifically focus on studying the smile as it is considered to be a rich, complex and sophisticated facial expression, formed through the synergistic action of emotions. According to Ekman [152], there are 18 different types of a smile, each of them corresponds to a specific situation and reflects a different type of emotion. Moreover, various studies show that there are differences in smiles between males and females, i.e. females tend to bear more expressive smiles than males. Furthermore, recent research indicates that females express emotions more accurately in both spontaneous and posed situations, e.g.[98, 163]. The literature review has been described in detail in Section 2.3. In this section, we specifically highlight the distinctiveness of our work in this area.

Based on the findings from such psychological studies, we examine the intensity and the duration of a smile in the hope of finding a distinction between the two sexes. Hence, in this chapter, we present an algorithm to measure gender solely based on the dynamics of the smile without resorting to appearance-based image analysis techniques. The dynamic framework we have developed for smile analysis has four key components. They are the spatial dynamics of the face based on geometric distances across the entire face, dynamic triangular areas of the mouth, the geometric flow across key areas of the face and statistically inspired intrinsic features which further analyse the spatial and area parameters. These purely dynamic features are then fed to a machine learning routine for classification, resulting in an algorithm for gender recognition.

Based on the hypothesis that there are differences in the smile of the genders, we designed an algorithm to measure the smile or the changes in the lip and eye areas, especially the dynamic intensity of those areas of the face, to find clues which can distinguish gender differences. Thus, our framework uses the dynamic changes that occur on the face, whereby time intervals are measured from the moment the neutral expression starts, to the peak of the smile. This research has been successfully published in the Visual Computer journal [41].

The application of this approach is to gain the machine the ability to a deeper understanding of human emotion where we prove that it contains hidden clues such as gender. Additionally, our proposed method will beneficial in case of a lake of texture data which our approach can overcome by simply analysing the dynamic feature of the expression.

6.1 The Computational Framework for Smile Dynamics

It has been hypothesised and evidenced by various psychological experiments that there exist differences in smiles between the two genders. To verify this computationally and at the same time to develop a tool for gender classification solely based on the smiles, we propose a framework which can track the dynamic variations in the face from neutral to the peak of a smile. Our framework is based upon four key components. They are (1) the spatial features which are based on dynamic geometric distances on the overall face, (2) the changes that occur in the area of the mouth, (3) the geometric flow around prominent parts of the face and (4) a set of intrinsic features based on the dynamic geometry of the face. Note, all of the dynamic features described here are intuitive extensions of the relevant physical experimentations and are based on the reported literature on facial emotions, especially on the dynamics of the smile, e.g. [97, 163].

Figure 6-1 presents a block diagram showing the key components of our framework for the analysis of the dynamics of a smile. The first step in our framework is to detect and track the face within a given video sequence. To do this, we have used a well-known Viola-Jones algorithm. It is based on Haar feature selection to create an integral image through the use of Ada-boost training and cascade classifiers [24]. The ability of this algorithm to robustly detect faces under different lighting conditions is well established, and we have also demonstrated this in previous work [40].

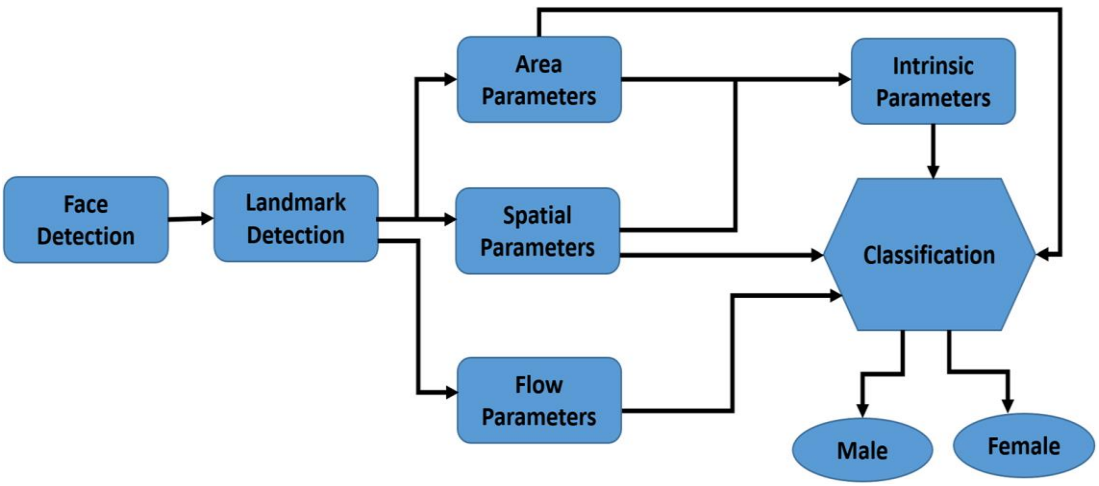


Figure 6-1 : Proposed framework.

The next step in our proposed framework involves automatic detection and tracking of a stable set of landmarks on the dynamic face. Automated Landmark detection is done using the CHEHRA model[158]. The algorithm has been trained to detect facial landmarks using in-the-wild datasets under various illumination, facial expressions and head pose. It is based on cascade linear regression for discriminative face alignment. This is done by applying the Incremental Parallel Cascade of Linear Regression (iPar-CLR) method. The tests we have carried using the CHEHRA model appear to be acceptable though we noticed it is likely to suffer when it comes to real-time applications. The algorithm has been utilised to detect 49 landmarks on the face, marked as $P_1 \ldots P_{49}$ as shown in Figure 6-2 (b) for the face shown in Figure 6-2 (a). Note, in addition to the 49 landmarks which CHEHRA detects, we also include the centres of the eyes as two additional landmark points, as shown in Figure 6-2, marked as P_{50} and P_{51}.

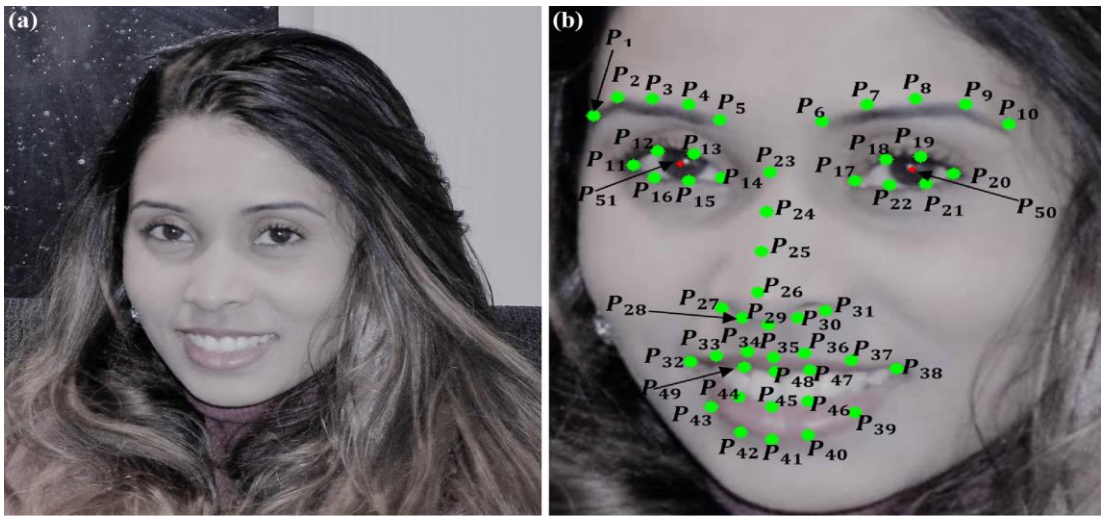


Figure 6-2: Forty-nine landmark detections using the CHEHRA model.

6.1.1 Dynamics of the Spatial Parameters

Based on some of the positions of the 49 landmarks we obtain through the CHEHRA model, we identify 6 dynamic geometric Euclidean distances across the face which are utilised to compute our dynamic spatial parameters. Further details of these spatial parameters are given in Table 6-1. The general form of a given spatial parameter can be computed using equation 6.1

$$\delta d_i = \frac{d_i}{N_i} + \sum_{n=1}^{t} \frac{d_i}{N_i} - \frac{d_{in}}{N_{in}}, \tag{6.1}$$

Where t is the total number of video frames corresponding to each $\frac{1}{10}th$ increment of the total time T for the smile, from neutral to the peak. Here N_i is the length of the nose, for a given video frame, computed as the distance between P_{23} and P_{26}. Thus, by dividing the spatial parameters by the length of the nose N_i, we normalise these parameters to the given dynamic facial image. It is noteworthy to point out that for a given smile, from neutral to the peak, we divide

the time it takes into ten partitions and therefore for each of the d_i we have 10 times d_i parameters which are fed to the machine learning. Hence, in our dynamic smile framework, we have a total of 60 dynamic spatial parameters.

Table 6-1 : Geometric distance and area.

Distances and area	**Description**	**Description by landmarks/area**
$\boldsymbol{d_1}$	Distance between mouth corner points	Distance (P_{32} , P_{38})
$\boldsymbol{d_2}$	Distance between upper and lower lip	Distance (P_{45} , P_{48})
$\boldsymbol{d_3}$	Distance between left mouth corner and left nose corner	Distance (P_{32} , P_{27})
$\boldsymbol{d_4}$	Distance between right mouth corner and right nose corner	Distance (P_{38} , P_{31})
$\boldsymbol{d_5}$	Distance between left mouth corner and left eye outer corner	Distance (P_{32} , P_{11})
$\boldsymbol{d_6}$	Distance between right mouth corner and right eye outer corner	Distance (P_{38} , P_{20})

Figure 6-3 shows the variation of δd_i across the 10-time partitions for a typical smile. As can be observed, there is a variation in each parameter as the smile proceeds from neutral to its peak.

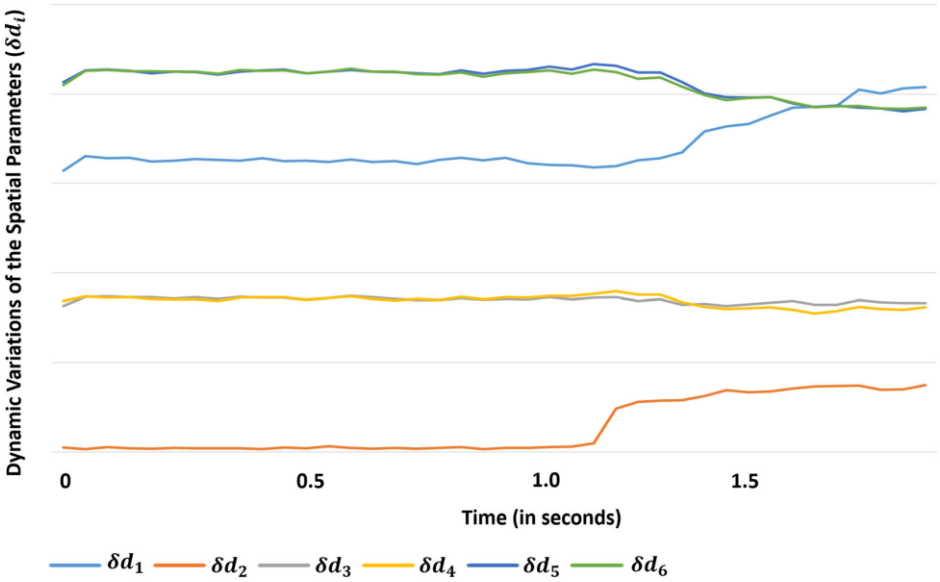


Figure 6-3: Variation in the dynamic spatial parameters δd_i, across the 10 partitions of time, for a typical smile, from neutral to the peak.

6.1.2 Dynamic Area Parameters on the Mouth

The second set of dynamic parameters concerns the mouth. Here we compute the changes in the area of 22 triangular regions that occupy the total area of the mouth. This is shown in Figure 6-4. Again, these areas are computed using the corresponding landmarks obtained from the CHEHRA model. The general form of how the changes in the mouth area are computed is described as,

$$\Delta_{area}^{i} = \sum_{n=1}^{22} \frac{\Delta_i}{\Delta N_i} \ , \tag{6.2}$$

and,

$$\delta\Delta_i = \sum_{n=1}^{t} \Delta_{area}^{i} \ , \tag{6.3}$$

where t is the total number of video frames corresponding to each $\frac{1}{10}th$ increment of the total time T for the smile, from neutral to the peak. Here ΔN_i is

the invariant triangle area determined by the landmarks defining the outer corners of the eyes and the tip of the nose, namely P_{11}, P_{20} and P_{26}. Again, we divide the total time of the smile, from neutral to peak, into ten partitions, and therefore we obtain 10 parameters from the $\delta\Delta_i$,though time, which is fed to the machine learning. Thus, in our dynamic smile framework, we have a total of 10 parameters which capture the dynamics of the mouth. For the purpose of illustration, in Figure 6-5 we show the distribution of areas of the triangular regions, Δ_i ,across a typical smile.

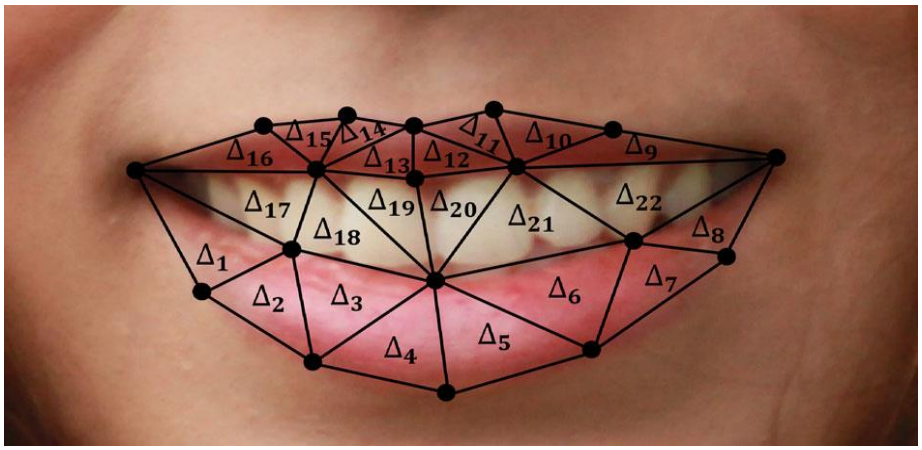


Figure 6-4 : Description of triangular mouth areas used to form the dynamic.

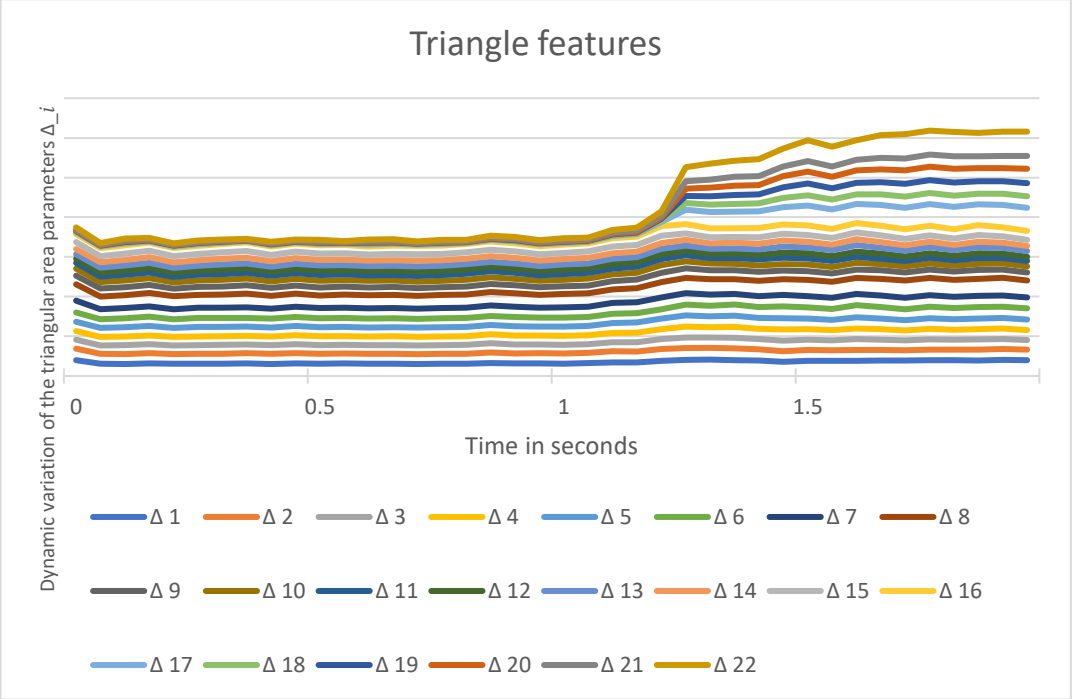


Figure 6-5 Variation in the dynamic area parameters i on the mouth, across the 10 partitions of time, for a typical smile, from neutral to the peak.

6.1.3 Dynamic Geometric Flow Parameters

The third component of our framework for smile dynamics is the computation of flow around the face during a smile. More specifically, we compute the flow around the mouth, cheeks and around the eyes. To do this, we have utilised the dense optical flow developed by Farnebäck [147]. It is a two-frame motion estimation algorithm in which quadratic polynomials are used to approximate the motion between two subsequent frames in order to approximate motion of neighbourhood pixels for both the frames. Using this algorithm, we are able to estimate the successive displacement of each of the landmarks during the smile.

Table 6-2 shows how the various landmarks and regions of the face are utilised to compute the optical flows around the face. The relevant facial regions

and landmarks are given in Figure 6-2 (b) and Figure 6-7 respectively. We also show the variations in the dynamic optical flows, δf_i, around the face for a typical smile in Figure 6-8.

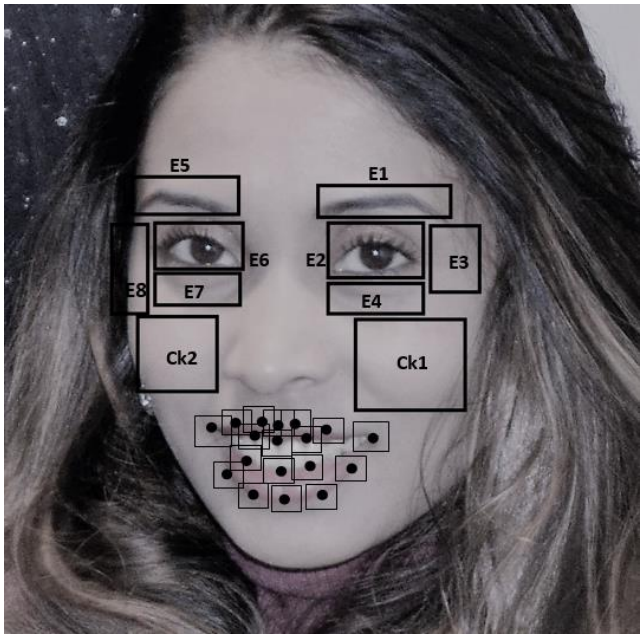


Figure 6-6 :Regions of the face identified for dynamic optical flow computation.

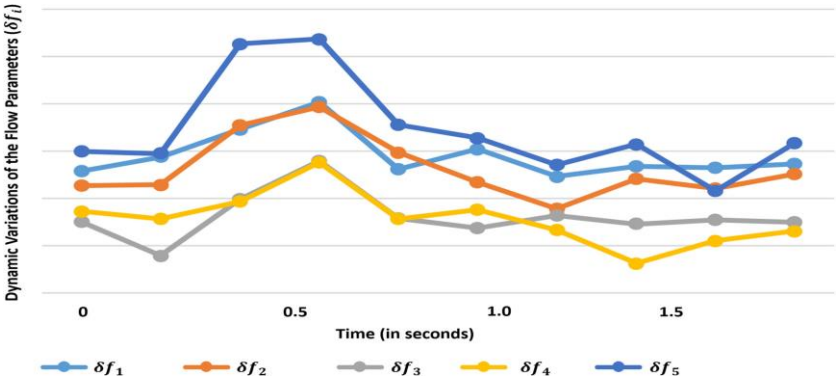


Figure 6-7 Variations in the dynamic optical flows around the face, for a typical smile, from neutral to the peak.

Table 6-2 : Description of how the optical flow parameters around the face are derived.

Optical flow	Description	Landmarks/ Regions
$\boldsymbol{\delta f_1}$	Mouth	Landmarks $P_{32}$to P_{49}
$\boldsymbol{\delta f_2}$	Left eye	f_6, f_7, f_8, f_9
$\boldsymbol{\delta f_3}$	Right eye	f_1, f_2, f_3, f_4
$\boldsymbol{\delta f_4}$	Left cheek	f_{10}
$\boldsymbol{\delta f_5}$	Right cheek	f_5

Note, the geometric flow, δf_i, for each of the regions is normalised upon computation by means of the corresponding flow around the invariant triangle area of the face determined by the landmarks defining the outer corners of the eyes and the tip of the nose, namely P_{11}, P_{20} and P_{26}. Again, each of the geometric flow parameters δf_i is computed across the 10-time interval through which the smile is measured, resulting in a total of 50 dynamic geometric flow parameters which are then fed to machine learning.

6.1.4 Intrinsic Dynamic Parameters

In addition to the spatial parameters, the area parameters and geometric flow parameters, we compute a family of intrinsic dynamic parameters on the face to further enhance the analysis of the dynamics of the smile. These intrinsic parameters are mainly based on the computation of the variations in the slopes and the growth rates of various features across the face. We identify these features as S_1, S_2, S_3, and S_4, details of which we describe as follows. The first

parameter family in this category relates to the computation of the overall slope variation around the mouth during a smile. To compute this, we use,

$$s_{1i} = \frac{N\sum_{n=1}^{N} P_{ix}P_{iy} - \sum_{n=1}^{N} P_{ix}\sum_{n=1}^{N} P_{iy}}{\sum_{n=1}^{N} P_{ix}^2 - \left(\sum_{n=1}^{N} P_{ix}\right)}, \quad (6.4)$$

where N is the number of video frames comprising the whole smile, from neutral to the peak, P_{ix} and P_{iy} are the Cartesian coordinate equivalents in the image space corresponding to the landmark point P_i. Hence, a total of 12 parameters are identified for the variations in slopes around mouth corresponding to the mouth landmarks P_{32} to P_{43}.

The second family of parameters, S_2, in this category corresponds to the growth rates across smile corresponding to the spatial parameters as well as area parameters on the mouth. The growth rates arising from the spatial parameters are defined as,

$$s_{2i(spatial)} = \sum_{n=1}^{N} \frac{\delta d_i^t - \delta d_i^{t+1}}{\delta d_i^t}, \quad (6.5)$$

and for the area parameters on the mouth are,

$$s_{2i(area)} = \sum_{n=1}^{N} \frac{\Delta_i^t - \Delta_i^{t+1}}{\Delta_i^t}, \quad (6.6)$$

where N is identified as the total number of frames in the video sequence of the smile while t to t +1 defines two successive video frames. In addition to the growth rates $s_{2i(area)}$, for each of the 22 triangular regions of the mouth, we also compute the total growth rate for the mouth, by using Eq. (6.6) along with the 22

triangular mouth area information. This means we have a total of 6 + 22 + 1 = 29 parameters of dynamic intrinsic type S_2.

The third family of parameters,S_3 , in this category we have identified is for both spatial parameters across the face and area parameters in the mouth. These are defined as compound growth rates given as,

$$s_{3i(spatial)} = \left(\frac{\delta d_i^{neutral}}{\delta d_i^{Peak}}\right)^{1/N} - 1\,, \tag{6.7}$$

and,

$$s_{3i(area)} = \left(\frac{\Delta_i^{neutral}}{\Delta_i^{Peak}}\right)^{1/N} - 1, \tag{6.8}$$

where N, like previously, is the total number of frames in the video sequence of the smile. The compound growth rate is measured simply using the neutral and peak of the smile. Again, like previously, in addition to the compound growth rates $s_{3i(area)}$ we also compute the compound growth for the entire mouth by means of the utilising the total area of the mouth. This implies that we obtain a total of 29 parameters of dynamic intrinsic type S_3 too.

For the final family of parameters, S_4, in this category, we compute the gradient orientation of the mouth based on the two mouth corner landmarks P_{32} and P_{38} which provides us with a line m passing δd_1, at the neutral and the peak of the smile. We then use,

$$s_{4i} = \sum_{t=1}^{T} \delta d_1^t - m^t, \tag{6.9}$$

to compute the rate of deviation of the mouth corners against the gradient m over the 10-time partitions where T is the total time from neutral to the peak of the smile. Similarly, we compute the gradient orientation of the mouth area based on the combined 22 triangular areas of the mouth between the neutral frame and the peak of the smile. These parameters provide us with a sense of the smoothness of the smile and forms an additional 10 + 10 = 20 parameter for machine learning. Table 6-3 provides a summary and brief description of various parameters associated with our computational framework for smile dynamics.

Table 6-3 : Parameter description for the computational framework.

Parameter	Description	Number of parameters
$\boldsymbol{\delta d_i}$	Spatial—involving 6 geometric distances across the face	60
$\boldsymbol{\delta \Delta_i}$	Mouth area—derived from the total area for the 22 parts of the mouth	10
$\boldsymbol{\delta f_i}$	Geometric flow around the mouth, eyes and cheeks	50
$\boldsymbol{s1_i}$	Slope measurements around mouth landmarks P_{32} to P_{43}	12
$\boldsymbol{s2_i}$	Growth rates of the spatial parameters and mouth areas	29
$\boldsymbol{s3_i}$	Compound growth rates of the spatial parameters and mouth areas	29
$\boldsymbol{s4_i}$	Gradient orientations for the mouth corners and the mouth area	20

6.2 Experiments

Once an appropriate framework for the analysis of the dynamics of the similes, as described above, is in place, we carried out a series of experiments to further analyse the pattern of smile and more importantly to look for clues of gender in the smile. For this purpose, we utilised two well-known datasets to carry out an initial set of experiments. We then utilised the same datasets to extract the parameters described in Table 6-3 and fed them to machine learning routines.

6.2.1 Datasets

We tested our approach on two main datasets namely, the CK+ and the MUG datasets. The CK+ dataset has a total of 83 subjects, consisting of 56 females and 27 males. The smile of each of the subjects went from the neutral expression to the peak of the smile. On the other hand, the MUG dataset contains a total of 26 subjects, consisting of 13 females and 13 males. The smile of each subject, in this case, went from the neutral expression through to the peak and finally returning to the neutral. Since our framework has been developed to analyse smiles from neutral to the peak, we modified the MUG dataset so as it only contained the relevant parts of the smile for each subject. In addition to this, for each smile, we also ensured that within the two datasets there contained an equal number of video frames. Thus, a total of 109 unique subjects were available to us for training and testing. For further details refer to Section 1.7.

6.2.2 Evaluation of landmark detection model

In this section, we evaluate the CHEHRA model as it is one of the key components of the proposed method. The CHEHRA model has been used in a variety of recent research on facial expression analysis and detection [164-166] and it proves to be stable in detecting facial landmarks and head pose.

For evaluating the CHEHRA model we performed a set of experiments for measuring the accuracy of it in detecting landmarks. As far as detecting landmarks goes the model 100% accuracy for both datasets as both datasets we have used. For checking for the accuracy of the model, we use the CK+ dataset as it has manually labelled landmarks to help evaluate the accuracy of the model. Thus, the evaluation is done by comparing the landmarks position of those manually coded and with that obtained from the CHEHRA model. For this experiment, to evaluate the CHEHRA model, we used a total of 82 videos.

In order to do the comparison, the CK+ dataset contains 68 landmarks and the CHEHRA model detects 49 on each face, and so we compared the corresponding landmarks to check for validity. Figure 6-8 shows the X-axis position of the manual coded (CK+) vs the results obtained for CHEHRA model (CH), as presented there is no significant difference between the two approaches as shown in the average standard deviation and high P-value. Furthermore,

Figure 6-9 shows the comparison of the Y-position of the landmarks which show there is no significant difference between the two models.

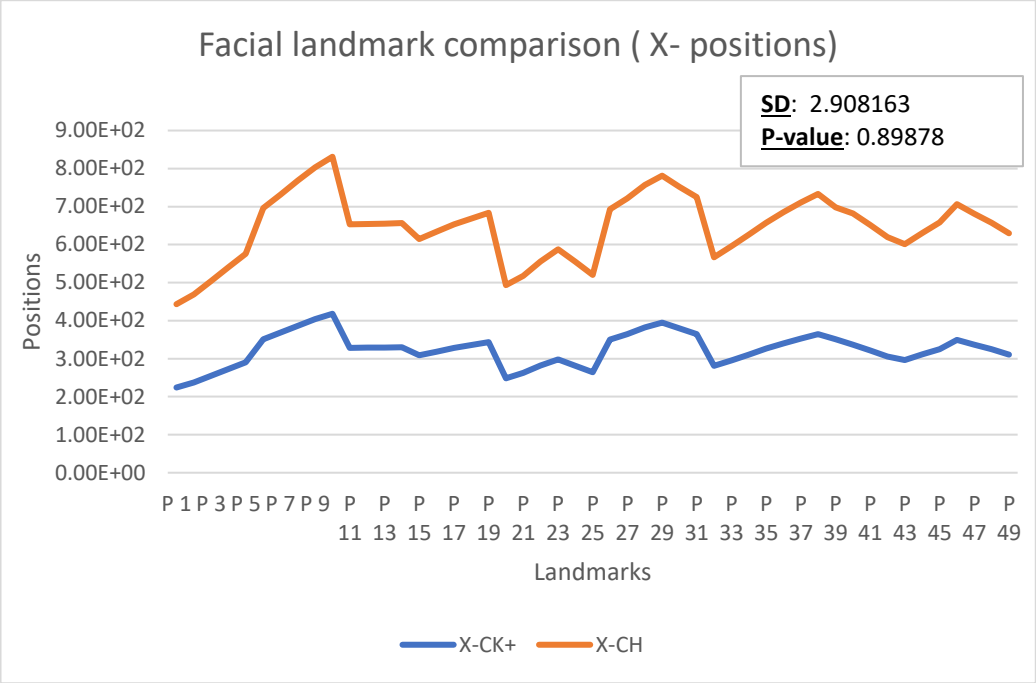


Figure 6-8: Comparison between the manual coded vs the CHEHRA model on the landmarks X- positions.

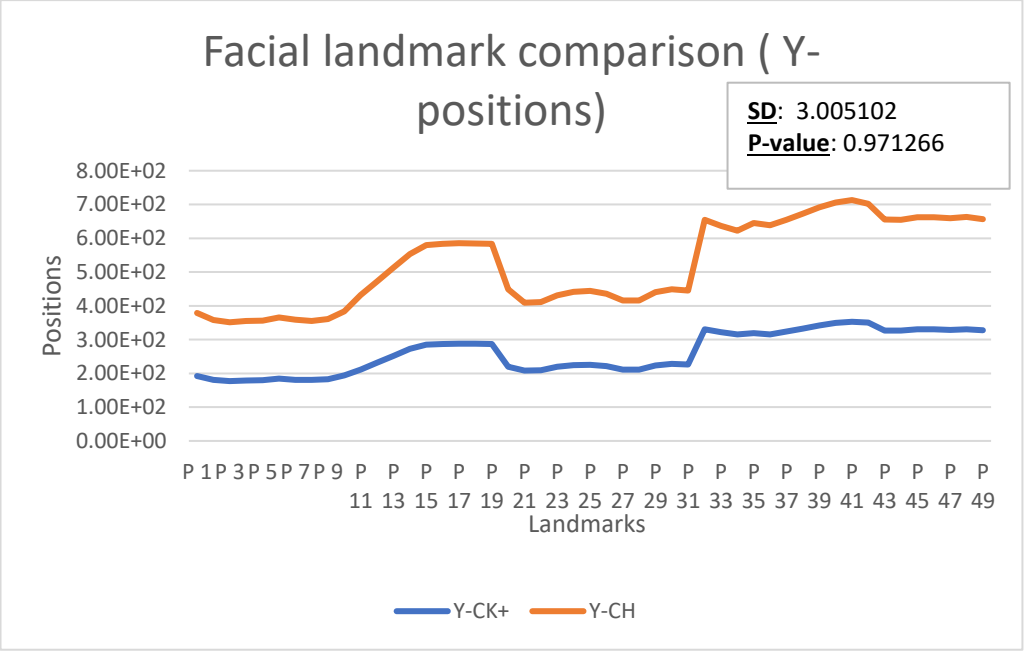


Figure 6-9: Comparison between the manual coded vs the CHEHRA model on the landmarks Y- positions.

6.2.3 Initial Experiments

Here we report an initial set of interesting experiments that we undertook to further understand the dynamics of smiles and to seek for clues of gender in smiles. In our first experiment, we tried a rather brute force approach to identify the areas of the face that contain most information of the smile that relates to the gender. Figure 6-10 shows some results based on the changes in the areas of the mouth region for 54 subjects (27 females and 27 males) in the CK+ database for the peak of the smile. As can be observed, there appears to be no significant difference between genders when one considers this simple form of analysis.

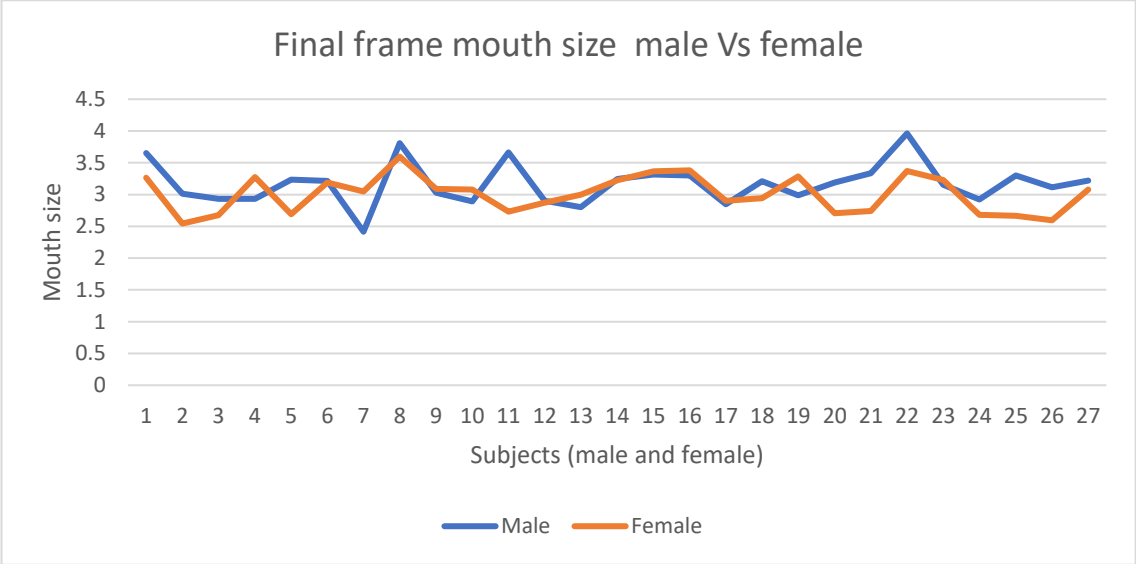


Figure 6-10: Variations in the area of the mouth at the peak if the smile for 54 subjects in CK+ dataset.

Next, we considered a computation relating to the product value for the areas for the upper lip and lower lip using the $\delta\Delta_i$ given in Equation (6.2) throughout the smile expression. This was done by multiplying each feature value for the changes in the mouth areas from a given video frame by the corresponding

values from the next frame, in order to obtain the product of the features (POF) through the smile expression, i.e.,

$$POF_i = \sum_{t=1}^{N} \Delta_i^t \Delta_i^{t+1} \quad (6.10)$$

where N is the number of video frames containing the smile expression from neutral to the peak.

Analysis of the POF data gave us some clues for gender difference between the smiles. Thus, by taking the average of each attribute, for both the genders, Figure 6-11 shows the POF for the females and the males in the CK+ dataset and Figure 6-12 shows the corresponding POF plots for the subjects in the MUG dataset. As can be inferred, from the average POF results shown in Figure 6-12, Figure 6-12, there appears to be a distinguishable difference in the smile of the males and females in terms of the POF computations.

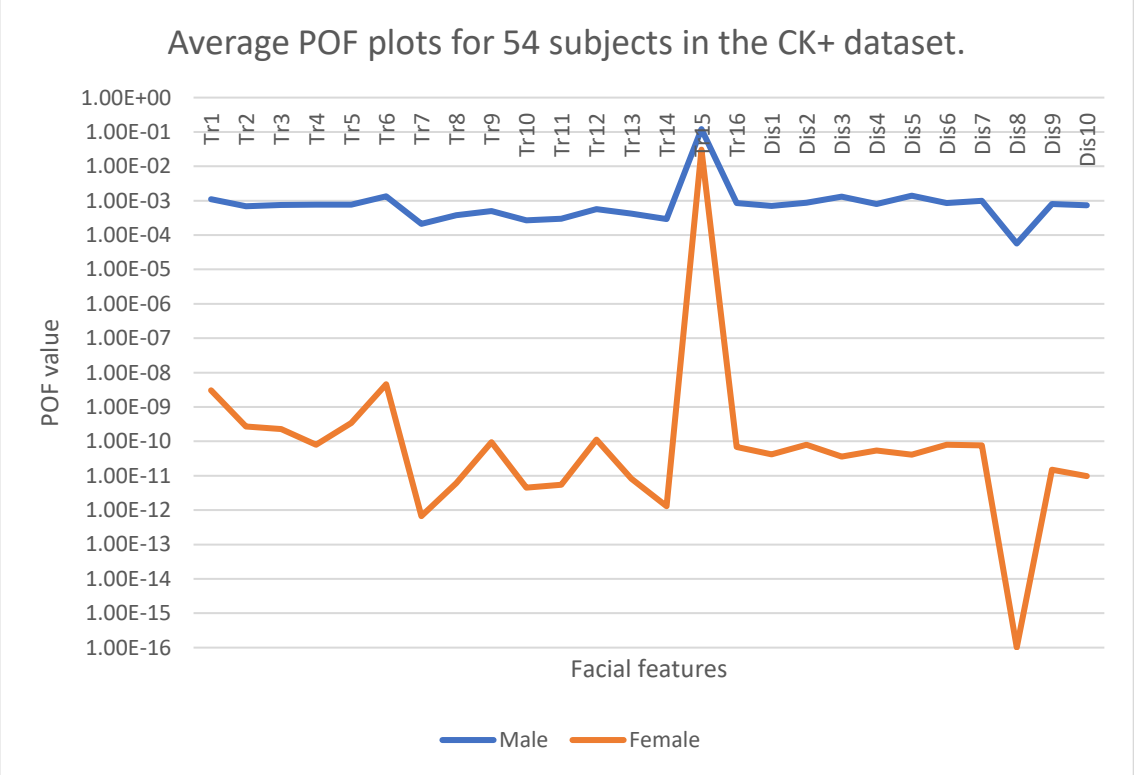


Figure 6-11: Average POF plots for 54 subjects in the CK+ dataset.

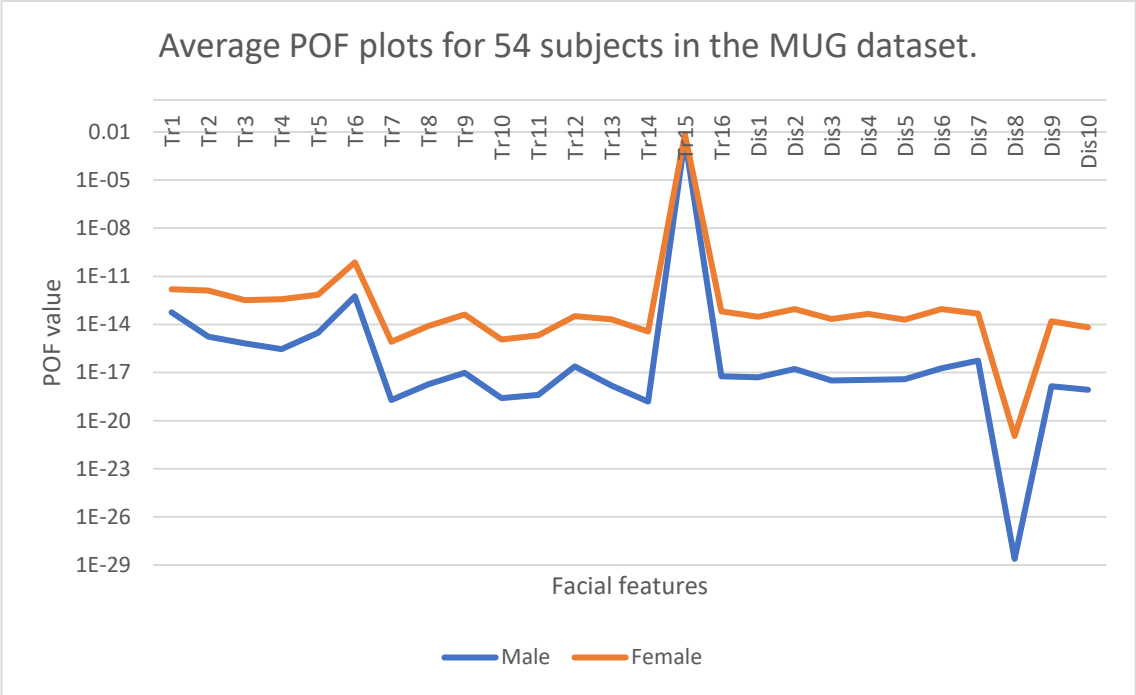


Figure 6-12: Average POF plots for 26 subjects in the MUG dataset.

Furthermore, from a first glance at these results, one might infer that males have a more intense smile than females which directly conflicts with the various psychological studies. However, that is indeed not the case. In fact, we note that in this experiment we computed the POF for each triangular features whose values are always less than 1. Additionally, for normalisation, we divided the POF values with the invariant area of the eyes-nose triangle. The result is a very small number, less than 1. Since the product of smaller numbers is smaller too, the POF values for females are smaller than that for males. Hence, it indeed confirms the smiles of females expand more through time in comparison to males.

Though rather simple, using the above approach, we were able to classify the data, through the median POF value computed from the mouth triangular attributes. This lends us a 60% correct classification for gender. That, however, is just slightly above chance and hence would not be considered an acceptable

method of classification. We then used all the features described in our computational framework for smile dynamics to train and test a machine learning classifier. With this knowledge, we constructed a suitable machine learning approach for classification. Following the procedure described in the block diagram shown in Figure 6-1, the following algorithmic (Algorithm 1) form further elaborates the computational framework we have developed for the automatic analysis of smile dynamics.

Algorithm 1 Gender from the dynamics of a smile.

M_j = Matrix of features in the smile
N_i = Sequence of video frames in the smile
H_k = Matrix of computed features from each video frame

while $k \neq N_i$ **do**
  Detect the face
  Detect the landmarks, P_k
  Compute mouth areas, $\triangle_k$
  Compute geometric distances, d_k
  Compute geometric flows, f_k
  Populate H_k
  k=k+1
end while

Compute spatial parameter values, $H_k \rightarrow \delta d_j$
Compute area parameter values, $H_k \rightarrow \delta \triangle_j$
Compute flow parameter values, $H_k \rightarrow \delta f_j$
Compute intrinsic parameter values, s_1, s_2, s_3 and s_4

Form data clusters $M_j \rightarrow D_n$, $n = 10$
Apply classifier to D_n
Output gender

6.2.4 Classification using Machine Learning

For our machine learning based classification, we have utilised two well-known classification algorithms namely, the support vector machine (SVM) and the k-nearest neighbour (KNN) as other algorithm reported very low classification

percentages because of the lower feature space we present in the proposed method, such algorithms include: Naïve Bayes classifier, decision trees and discriminant analysis. First, we tried to use PCA as a pre-step before applying SVM. The results indicated this approach yields a very low classification rate. This is probably due to the fact that PCA reduces the number of features, while at the same time eliminating some distinguishing features. Second, we used SVM on its own, without the PCA. We had a mild improvement in the classification rate of 69%.

Finally, we used the k-NN algorithm which is a nonparametric method used for classification and regression [167]. The output of the k-NN algorithm is a class relationship. The object can be assigned a class by k nearest neighbours where k is a positive integer. We utilised all the 210 features described in Table 6-3 to train our classifier. Additionally, we used a ten-fold cross validation scheme to validate our k-NN classifier. The results were tested on several distance functions namely, Euclidean, Cosine, Minkowsky and Correlation. In Table 6-4, we report the best results we have obtained using the k-NN classifier.

Table 6-4 : Results using the k-NN classification.

	CK+	**MUG**
k-NN distance	Correlation	Cosine
k value	3	14
Classification	78%	86%

6.3 Conclusions

In this research, we are concerned with the identification of gender from the dynamic behaviour of the face. In this sense, we wanted to answer the crucial question of whether gender is encoded in the dynamics of a person's smile. To do this, we have developed a computational framework which can analyse the dynamic variations of the face from the neutral pose to the peak of the smile. Our framework is based upon four key components. They are the spatial features which are based on dynamic geometric distances on the overall face, the changes that occur in the area of the mouth, the geometric flow around some of the prominent parts of the face and a set of intrinsic features based on the dynamic geometry of the face. This dynamic framework enables us to compute 210 unique features which can then be fed to a k-NN classifier for gender recognition.

We ran our experiments on a total of 109 subjects (69 females and 40 males) from two datasets, namely the CK+ and the MUG datasets. Firstly, our results do agree with that of various psychological studies, indicating that females are more expressive in their smiles. For example, this became evident to us by simply looking at the changes in the lip area during a smile in which the lip area of female subjects expands more in comparison with the male subjects. Further, and more importantly, using machine learning approaches, we can also classify gender from smiles. In particular, by means of the standard k-NN algorithm, we are able to obtain a classification rate of up to 86%, purely based on the dynamics of smiles.

We understand from the presently available literature that some of the recent work carried in gender classification can achieve over 90% recognition rates using hybrid models with a combination of geometric and appearance features which are both static and dynamic. This is particularly clear from the work presented in [101]. It is, however, noteworthy that our work is geared to study the gender classification rates purely based on the dynamics of a smile. In fact, some of the results reported in [101] indicate that using their chosen dynamic smile features they obtain a classification rate of 60%, whereas using the smile dynamics framework we have proposed, we are able to obtain a higher gender classification rate of over 75%. There is also an added advantage of using the dynamic features, as opposed to static images, for gender identification since it presents with the opportunity to infer gender from certain parts of the face such as the mouth and the eyes areas.

Going forward into the future, there are a number of directions in which this work can be further taken forward. It will be useful to see if it would be possible to enhance the classification rates using other correlation and sophisticated statistical analysis techniques. In this chapter, we have only used simple machine classification techniques such as SVM and K-NN, since our prime aim here was to demonstrate the power of smile dynamics in gender identification. We believe the utilisation of sophisticated machine learning techniques will further improve the results. We also believe this will be the case if novel machine learning techniques such as convolutional neural networks based deep learning (e.g.[97]) can be adapted to the problem at hand. However, having said that, we must also highlight the fact that such sophisticated machine

learning techniques usually require sufficient and significant training data which, as far as smiles are concerned, are scarce at present.

In addition to this, the results could be further tested and validated on other datasets. One deficiency of this present study is that we did not look deeply into the gender variation between posed and spontaneous smiles. We believe our framework has merit in providing much room for such detailed analysis to seek gender differences between the two types of smiles. Additionally, aside from the expression of a smile, other basic emotional attributes such as surprise, fear, anger and disgust can be studied in detail to look for cues to enhance gender recognition from facial expressions in general. We believe the framework we have presented in this chapter can easily be adapted to undertake such studies.

Finally, as we show that the gender is encoded within the dynamic characteristics of the smile expressions. Our aim in the next chapter is to show the possibility of using the dynamic characteristics of the smile as a biometric to identify the human identity.

7 Towards the Development of an Emotional Biometric

Biometric techniques can be defined as "*an automated method of verifying or recognizing the identity of a human based on a physiological or behavioural characteristic*" [7, 8]. The revolution of computer power and vision has enhanced biometric techniques over the years, as these techniques show the capability of recognising an individual's identity based on their physiological or behavioural characteristics. Additionally, biometric systems are considered as being anonymous recognition since they do not require any identification data such as age, gender, profession, residence or nationality [4].

Traditional methods of authentication techniques, which include passwords, PINs, smart cards, tokens, keys and other methods, show that they can be manipulated, stolen, misplaced, forgotten or duplicated. However, human physiological or behavioural characteristics show higher complexity and security levels compared to the traditional method where they cannot be misplaced, forgotten, stolen or forged. Physiological characteristics of biometric-based technologies include models such as the face, fingerprints, hand geometry, iris, retina, ear and voice. Behavioural characteristics include models such as gait, signature and keystroke dynamics. All these models show their ability to be used as biometrics to identify human identity with a high accuracy rate.

In this research, we investigate a new behaviour biometric model called facial expression biometric (FEB), which can be identified as the authentication process done by examining the way the person behaves and expresses emotions. The research in FEB is very limited and there are a couple of psychological and computational studies that try to identify the FEB and use it as a biometric technique (refer to Section 1.4.4 for more details).

Our aim of this research is to prove human identity is encoded within the dynamic characteristics of the facial expressions (smile expression). Furthermore, to identify the dynamic aspect of the smile expression where we study the stabilities of the different smiles for the same subject and the similarity aspect to other subjects. We believe that this type of biometric is very hard to be manipulated since they are controlled by the emotional state and muscle structure.

The application of this research is to identify another aspect of using the facial expressions in biometric filed. Furthermore, enhance the security within the state-of-the-art face recognition techniques by adding second level authentication using the FEB. Finally, it can be used in a different variety of application such as identity verification, fraud detection and surveillance.

This chapter is organized as follow. Section 7.1 shows an initial experiment which describes the using of PCA and CNN to identify the FEB. Section 7.2, discuss the result of the machine learning, problem and the motivation behind the proposed method. Section 7.3, the proposed method. Section 7.3 result and 7.4 the conclusions.

7.1 FEB using PCA and CNN

Using a machine-learning experiment explained in Appendix D, both CNN and PCA have a very high recognition rate (99%). Although a promising output of the machine-learning technique, it does not show the analytical part of the process in which we try to prove how many similarities there are among them and if the smile dynamic alone can be used to identify human identity. These

techniques are considered as a black box which processes the images and outputs them as a face recognition problem.

The results indicate that smile expression does not affect face recognition; this is due to a large number of features used. On the other hand, we notice that these large numbers of features contain parts that are not related to the smile dynamic. So, we designed a system to measure the dynamic features of the smile without taking into consideration any texture data, in order to study the possibility of recognising human identity using only the dynamic characteristics of the smile.

7.2 The Proposed Method

In this research, we designed a system to measure smile dynamic. Our system tries to analyse different features of human smiles that we try to prove are stable over time and show the smile dynamic can be used as a biometric. Our system consists of two main parts: smile intervals and smile analysis. Figure 7-1 shows the proposed system flow.

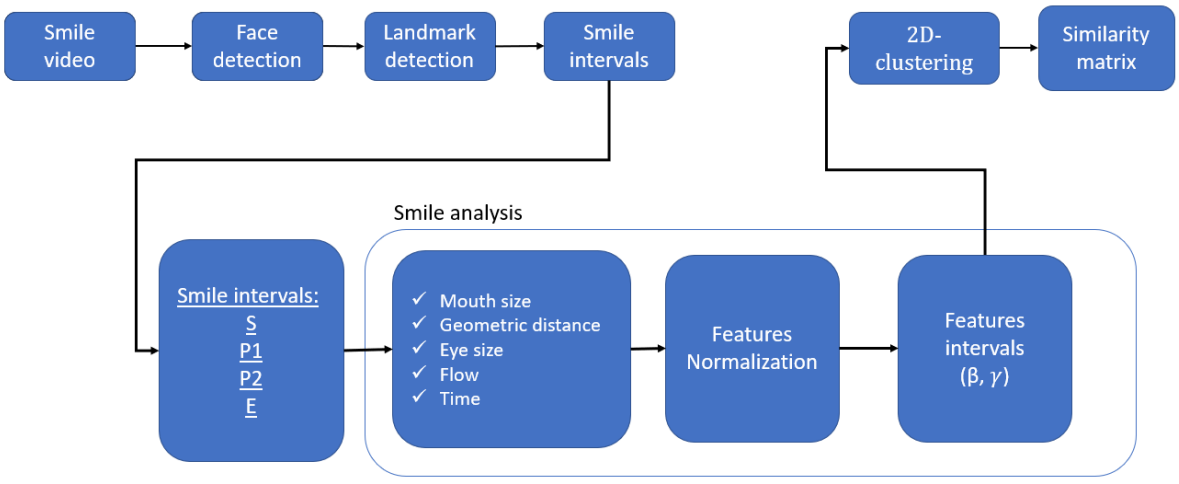


Figure 7-1: Proposed framework.

Face detection is done by using the Viola-Jones algorithm. Automated landmark detection was done using the CHEHRA model [158]. Figure 7-2 shows the 49 landmarks detected when applying the CHEHRA model. The evaluation of this model is done in Section 6.2.2.

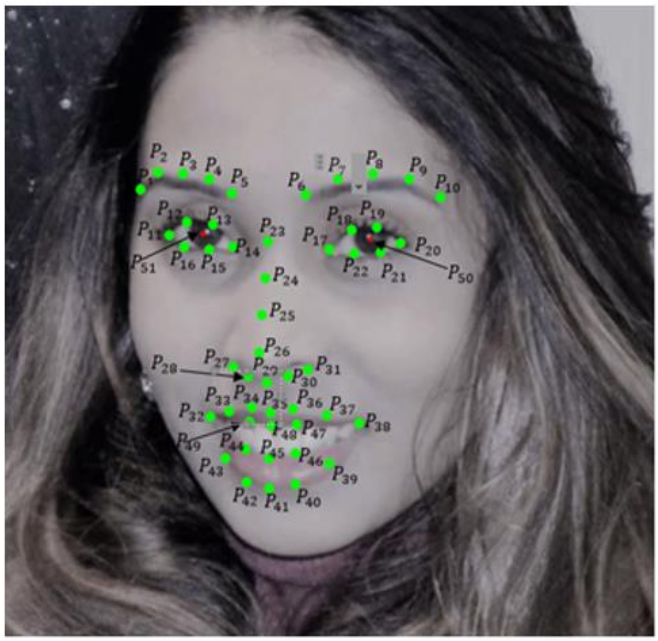


Figure 7-2: CHEHRA landmark detection.

7.2.1 Smile intervals

Smile interval represents the time at which the smile starts to reach the peak (onset) and stay there(peek), then return to the neutral expression (offset). Smile interval consists of four main points: identify the start of the smile (S), reaching the peak of the smile which is identified by the interval (P1, P2) and return to neutral expressions (E). Figure 7-3 shows an example of mouth size changes through the smile and identifies the smile intervals.

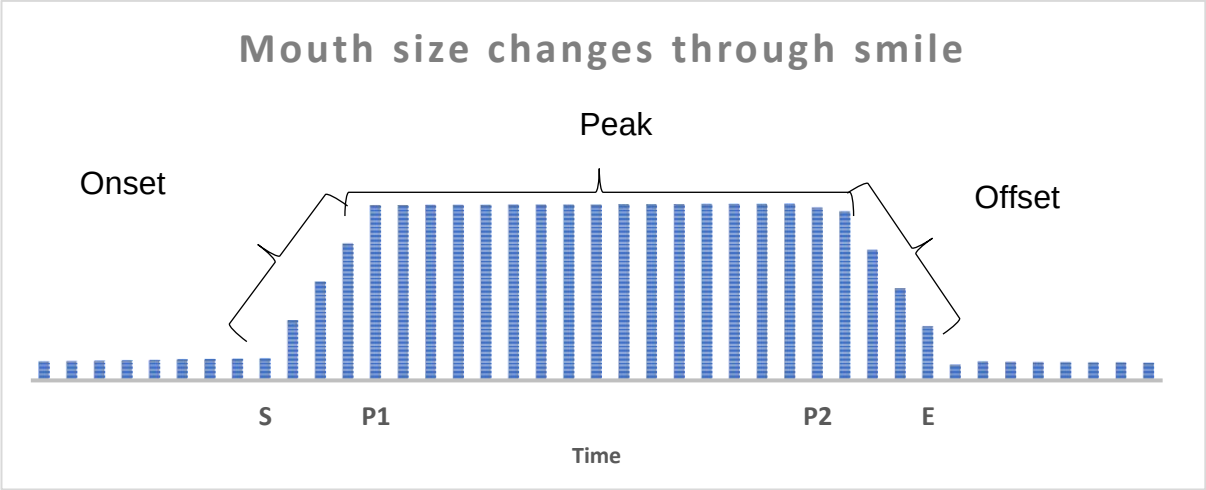


Figure 7-3: Mouth size change through smile expression with smile interval identified.

The purpose of computing smile interval is for more accurate analysis of the smile and to test if the time can be used as a distinguishing feature. Additionally, MUG dataset videos were edited by humans where each video has a lot of unrelated parts of the smile which include additional frames and a longer video. Figure 7-3 shows neutral face from the start of the video to the start of the smile (S), which has 20% of the smile video, in which the face has no expression, as well as the end period which includes the smile to the end of the video which covers an additional 19%. Our analysis shows the smile occurs with an average of 1.7 seconds using our proposed smile intervals.

Using the landmark, we reconstruct the mouth using triangle features as shown in Figure 7-4(a), using Euclidean distance to compute the distance between the landmark to contract the edge of the triangle then using the HERON's formula [168] to calculate the tringle area. HERON's formula allows the calculation of triangle area when given its three side lengths. Each triangle feature is then summed up to get the total mouth size ($\boldsymbol{m_1}$). This method is also used to

compute the eye area ($\boldsymbol{e_1}, \boldsymbol{e_2}$) as shown in Figure 7-4(b). Figure 7-1 shows a set of geometric distances used in both the smile interval computation and smile analysis.

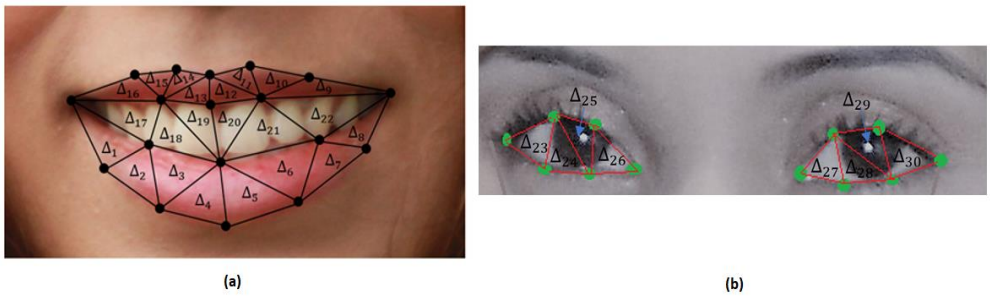


Figure 7-4: (a) Mouth dynamics using triangle feature, (b) Eye dynamics using the triangle feature.

Table 7-1: Dynamic features.

Distances and area	**Description**	**Description by landmarks/area**
$\boldsymbol{d_1}$	Distance between mouth corner points	Distance (P_{32} , P_{38})
$\boldsymbol{m_1}$	Total mouth area formulated by the triangle in figure (7-4)	$\sum_{i=1}^{22} \Delta(i) = \Delta_1 + \Delta_2 + \cdots . + \Delta_{22}$
$\boldsymbol{LL_1}$	Lower lip area formulated by the triangle in figure (7-4)	$\sum_{i=1}^{8} \Delta(i) = \Delta_1 + \Delta_2 + \cdots . + \Delta_8$
$\boldsymbol{UL_1}$	Upper lip area formulated by the triangle in figure (7-4)	$\sum_{i=9}^{16} \Delta(i) = \Delta_9 + \Delta_{10} + \cdots . + \Delta_{16}$
$\boldsymbol{e_1}$	Total left eye area formulated by the triangle	$\sum_{i=23}^{26} \Delta(i) = \Delta_{23} + \Delta_{24} + \cdots . + \Delta_{26}$
$\boldsymbol{e_2}$	Total right area formulated by the triangle	$\sum_{i=27}^{30} \Delta(i) = \Delta_{27} + \Delta_{28} + \cdots . + \Delta_{30}$

We designed an algorithm to compute the smile intervals using Table 7-1 ($\boldsymbol{d_1}$, $\boldsymbol{m_1}$). The following pseudo-code shows the algorithm used to identify smile intervals where we compute the $\boldsymbol{d_1}$ changes to the neutral frame $\boldsymbol{dN}$ where this indicates how much in percentages the distance changes through the smile expression as shown in equation 7.1. Compared to the threshold value $\boldsymbol{\varepsilon}$ we can identify the smile start point **S** as shown in equation 7.2. Using experimentation, we found the threshold value $\boldsymbol{\varepsilon}$ is set to 1.1.

$$\boldsymbol{dN_{1,i}} = {\boldsymbol{d_{1,i}}}/{\boldsymbol{d_{1,1}}}, \tag{7.1}$$

$$if\ \left(\boldsymbol{dN_{1,i}} > \varepsilon\right)\ , \tag{7.2}$$

$$S = i\ ,$$

where $\boldsymbol{dN_{1,i}}$ is the distance changes to neutral face which is the first frame in the smile video $\boldsymbol{d_{1,1}}$ we identify the start of the smile S by comparing it to ε which is set to (1.1) to identify the peak intervals (p1, p2). We first compute the average of the mouth size m_1 from the smile start to the end of the video.

ALGORITHM: SMILE INTERVAL DETECTION

```
// N → Number of frames
// Peak-AVG → Peak average
// Mouth-Size → m1
------- Compute Start Point --------------

for I = 2 to N
      if (dN1,i > 1.1) // 1.1 denote the (ε)
            S=I      // Denote the Start of the smile at frame (I)
            Break for
end for
-------- Compute average -------------

Peak-AVG = 0          // Denote computing Peak interval average

for j = S to N
      Peak-AVG =peak-AVG + Mouth-Size(j)
End for

Peak-AVG= Peak-AVG/(N-S)

-------- Compute peak intervals -------------

for h = S to N
      if (Mouth-Size(h) >= Peak-AVG)
            P1 = h            // Denote the Peak start at frame(h)
            Break for
End for

for f= N to P1
      if (Mouth-Size(f) >= Peak-AVG)
            P2 = f            // Denote the Peak end at frame(f)
            Break for
End for

-------- Compute end of smile --------------

for g = p2 to N
      if (dN1,g < 1)
            E = g           // Denote the end of the smile at
frame(g)
            Break for
End for
```

To detect the peak interval (P1, P2), first, we compute the "*peak average*" as we assume after locating the smile, start point peak will be next. Computing peak average is done on the mouth area m_1, where we compute the average from the frame S to the end of the smile video. Computing the peak average will create a virtual threshold line that will identify the peak interval. Identify (P1, P2)

is done by scanning the $\boldsymbol{m_1}$ value through the smile from the start point to the end of the video and comparing it to the peak average. P1 is identified by checking the mouth size at each frame to the peak average which identified the P1 from the start of the smile to the end of the video. P2 is identified by checking the mouth size at each frame to the peak average from the end of the video to P1. Identify the end of smile E is done by comparing the $\boldsymbol{d_1}$ changes to the neutral frame ($\boldsymbol{dN}$) from the P2 to the end of the video if it is less than the 1 that is indicated the end of the smile. We set the end threshold value to 1 because it indicates that the $\boldsymbol{d_1}$ is fully returned to the neutral expression.

The design algorithm for detecting the smile interval is based on the experiment done on the MUG dataset where we tried different features to identify the smile interval. To test our approach, we manually coded the smile interval in the MUG dataset where we identified the frame number that indicates each smile interval point (S, P1, P2, E).

To identify the smile interval accuracy, we manually coded the MUG dataset with smile interval attributes and comparing it to the proposed method. As a shown, we gain 95% correct time interval.

7.2.2 Smile Analysis

After identifying the smile intervals, we measured a set of features to identify the smile dynamic. We analysed the smile within the time intervals. Smile analysis contains different features such as:

1. Mouth corner distance changes
2. Mouth size changes
3. Upper and lower lips changes
4. Eye area changes
5. Flow value

6. Time
7. Facial features order of activation

The selection of these features is based on the landmark detection model and the dynamic characteristics these features have through smile expression. As shown in Table 7-1, where each feature is computed in different ways and used for different purposes, we use $\boldsymbol{d_1}$ and $\boldsymbol{m_1}$ to identify the smile intervals. We studied both the dynamic characteristics of $\boldsymbol{d_1}$, which represent the distance between the mouth corners, mouth area which represents the dynamic change of the upper, lower lips and between the lips area. Eye area dynamics represents the eye area changes through smile expressions which capture changes in the eye through the smile expression.

Flow value represents the displacement of pixels caused by the smile formulation in specific regions of the face where we use a dense optical flow algorithm to measure these displacements presented by [147]. Where using landmarks we allocate a set of regions of interest that cover eyes, cheeks and mouth as shown in Figure 7-5 Flow computational was used for two purposes: first, to compute overall displacement of the face through the smile; second, to measure order of activation for each facial feature.

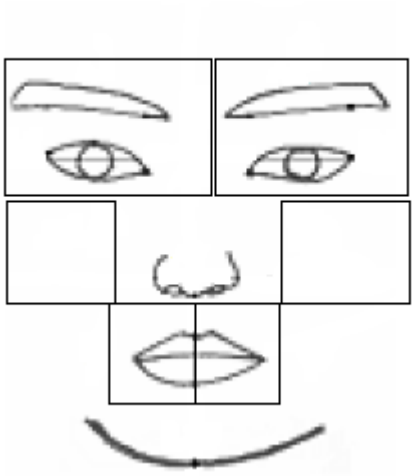


Figure 7-5: Regions of interest.

Computing overall displacement is done by summing up all displacement value in each region of interest in terms of smile intervals. This is done by computing flow value using dense optical flow (for more details refer to Appendix A). Computing order of activation for each facial feature is done by comparing the flow value for each region and taking them in order from the area with the maximum flow to the lowest.

Lastly, we measure the smile time presented by smile intervals. Our study of the time is to see if the same subject with a different smile has the stability of producing the same intervals on different occasions.

After calculating each feature, we normalise it by dividing with the triangle contracted by the tip of the nose and the outer corner of the eyes. The normalisation process unifies the finding for each subject and eliminates distance factor to the camera. Normalising the features is done by applying the following equation:

$$FE_n = {}^{FE_{n,i}}\Big/_{\Delta_{11,20,29}}, \quad (7.3)$$

where FE_N is the feature for ID n and i represents the frame number and ($\Delta_{11,20,29}$) is the triangle contracted by the eyes outer corner (p_{11}, p_{22}), points and tip of the nose (p_{29}).

We measure dynamic features and we use different machine-learning techniques to identify the face, landmark and the flow of each feature, which all contain an error rate in their computations which will reflect on smile dynamic features variation within the same subject expressing the same smile. Additionally, since we have more than one smile for each subject we predict there will be small variations in computing the dynamics of the smile since these features have dynamic characteristics. To overcome this problem, we compute the features intervals represented by (β, γ) where β represents the high features interval and γ represents the low interval of the features. Combining (β, γ) will create a virtual cluster that represents the average of smile dynamic for a specific subject.

The computation of both intervals is done by first computing the (μ, std) where μ_n is the minimum average for smiles for the same subject and std_n is the maximum standard deviation for different smiles for the same subject. Since our dataset contains 20 subjects with an average of three smiles each, computing the (β, γ) will show different aspects of consistency and stability within the same subject. Furthermore, it will be used for classification with other smiles to signify the uniqueness of the smile. Computing (β, γ) was done on the MUG dataset with an average of three smiles per subject using the following equations:

$$\mu_n = \frac{1}{s_n}\sum_{d=1}^{s_n}\left\{\frac{1}{m}\sum_{i=1}^{m} AVG\_FE_{n,i}\right\}, \quad (7.4)$$

$$std_n = \text{Max}\left\{\sqrt{\frac{1}{z}\sum_{j=1}^{z}\left(FE_{n,j} - mean_n\right)^2}\right\}_1^{s_n} \quad (7.5)$$

$$\beta_n = \mu_n + \quad std_n\,, \quad (7.6)$$

$$\gamma_n = \mu_n - \quad std_n\,, \quad (7.7)$$

where μ_n is the average of features n computed to over s number of smiles for a specific subject where each smile has specific averages $AVG\_FE_{n,i}$ computed within each smile interval m. std_n represents the maximum of standards division computed on different smiles for the same subject n over the number of smiles s. β_n represents the high interval for features n which is computed by adding the average of multiple smiles μ_n to the maximum of standards deviation value std_n. γ_n represents the low interval for features n which is computed by subtracting the average of multiple smiles μ_n to the maximum of standards deviation value std_n.

7.2.3 Computing Similarities

For classification, we use 2D-clustering where we use both (β, γ) to identify the boundary for each feature cluster related to the specific subject. These clusters represent different features in a 2D space with a boundary set by computing the (β, γ). To compute the smile similarity using the proposed features we compare the dynamics of each feature if it is within the boundary (β, γ) which is represented as a score and related to the specific subject. Summing up all the smile scores for different factures represents the smile similarity for different

subjects. For visualising purposes, Figure 7-6 shows two clusters for $\boldsymbol{d_1}$ and $\boldsymbol{m_1}$ for one subject in the MUG dataset. As shown, each feature has a (β, γ)

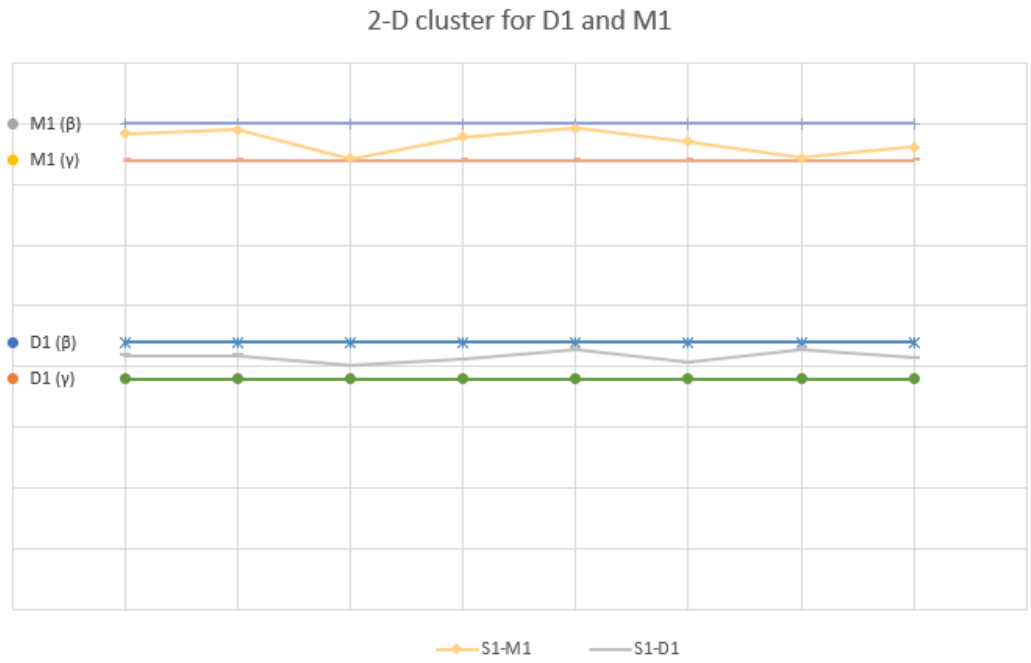


Figure 7-6: Visualising 2D clustering.

The process of creating a similarity matrix is done by comparing the (β, γ) and then using the features with smile intervals to identify the subject identity. Different intervals have different ways of computing (β, γ) where the start interval is computed over three frames around the S point, peak is computed through the number of frames within (p1, p2) and end is computed with three frames around the E point.

7.3 Results

Our analysis was done on the MUG dataset where we used 20 subjects, where each subject had an average of three smiles. For testing our smile interval detection, we compare our proposed method to the manual set of smile intervals

which we add to the MUG dataset. Figure 7-7 shows the proposed method detected intervals (S-P, P1-P, P2-P, E-P) versus the manual time intervals (S, P1, P2, E) for 10 subjects. Our algorithm has a 95% approval with the manual coded.

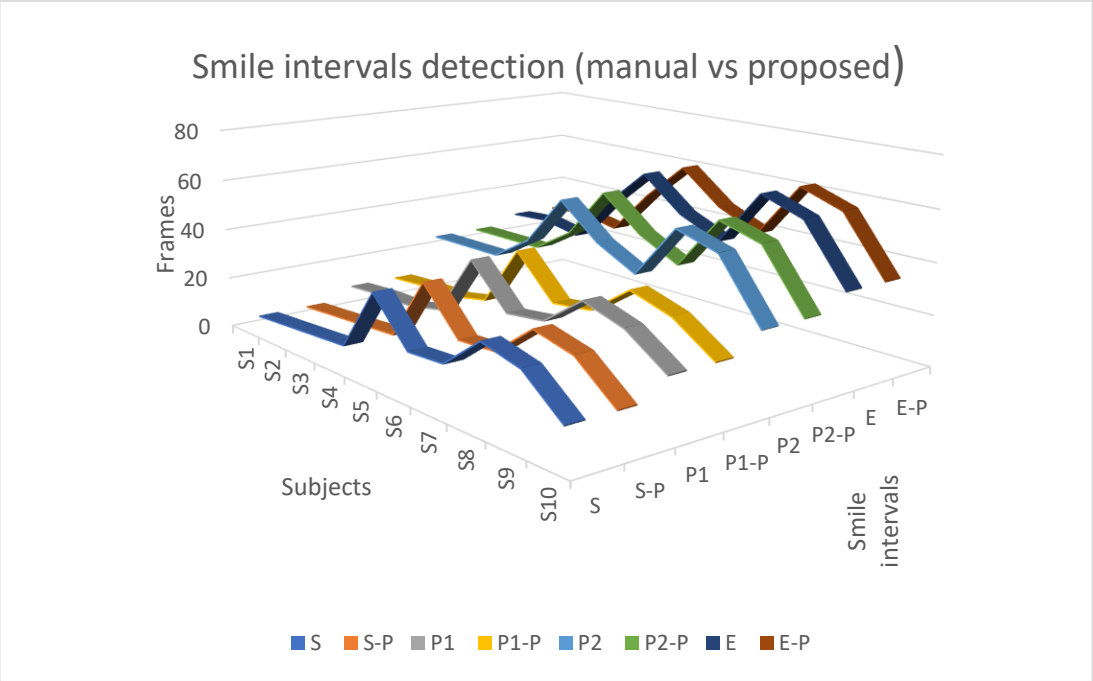


Figure 7-7: Smile interval - manual vs proposed.

To test our proposed dynamic features, we first examine how many similarities each feature contains. For finding similarities we use our 2D-clustering technique to examine how much each feature is similar or unique for all the smiles in the MUG dataset. In the first experiment, we compute similarity for each time interval (S, P1-P2, E). Figure 7-8 shows the similarity percentages for the dynamic facial features in the start of the smile.

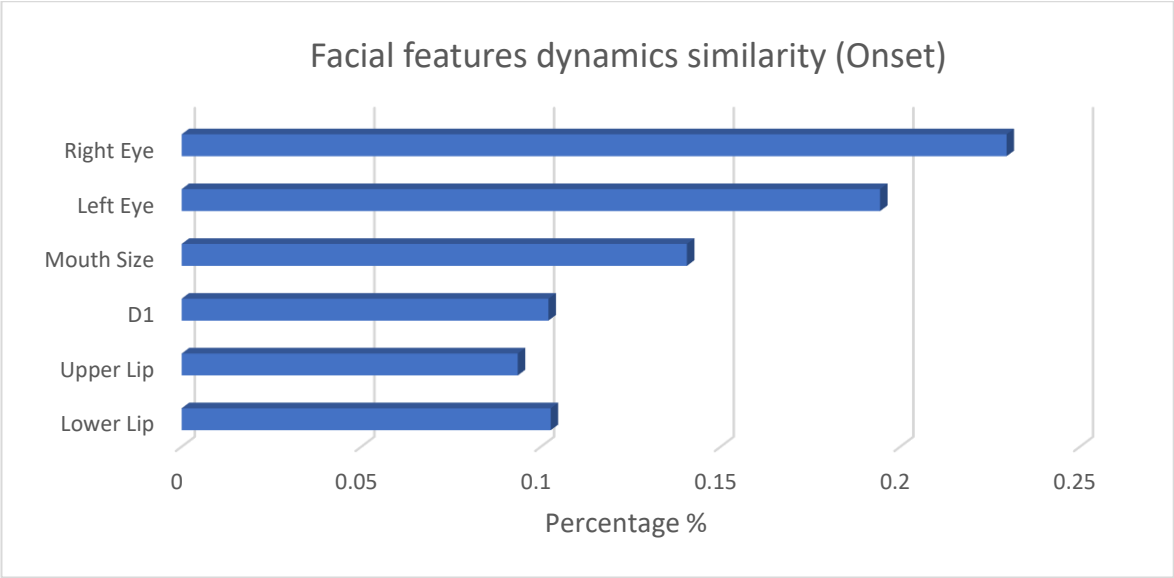


Figure 7-8: Dynamic features similarity at the onset interval.

Figure 7-9 shows the facial features similarity at the peak of the smile which shows the uniqueness of each facial feature in peak interval. As an example, as shown the lower lip percentages shows 10% similarity which implies that three smiles' have similar lower lips dynamic in the peak period. Figure 7-10 shows the facial features similarity at the end of the smile.

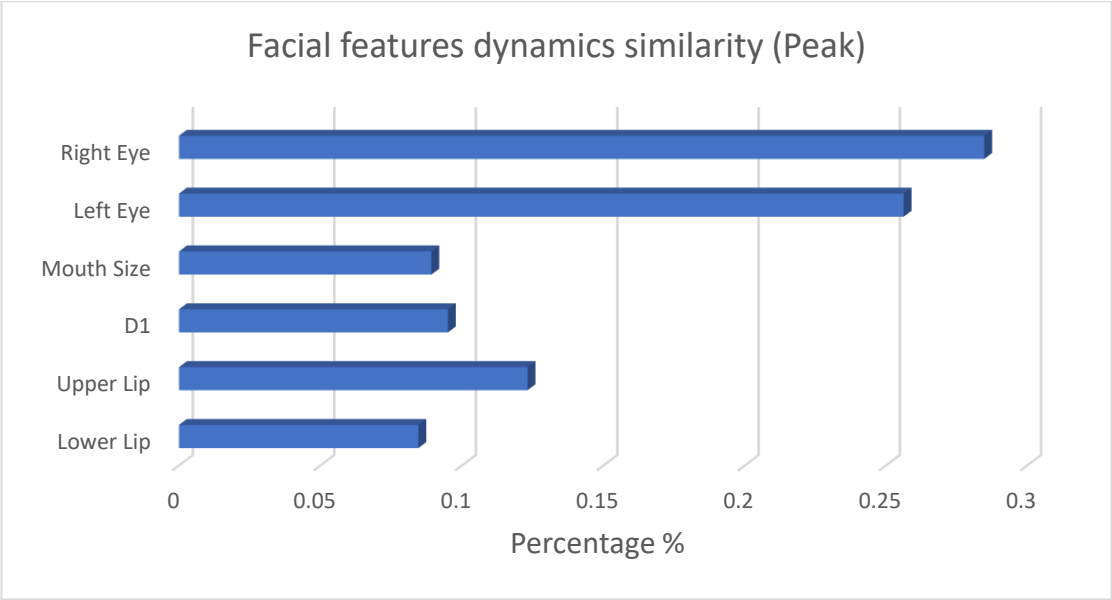


Figure 7-9: Similarity of the dynamic features at the peak of the smile.

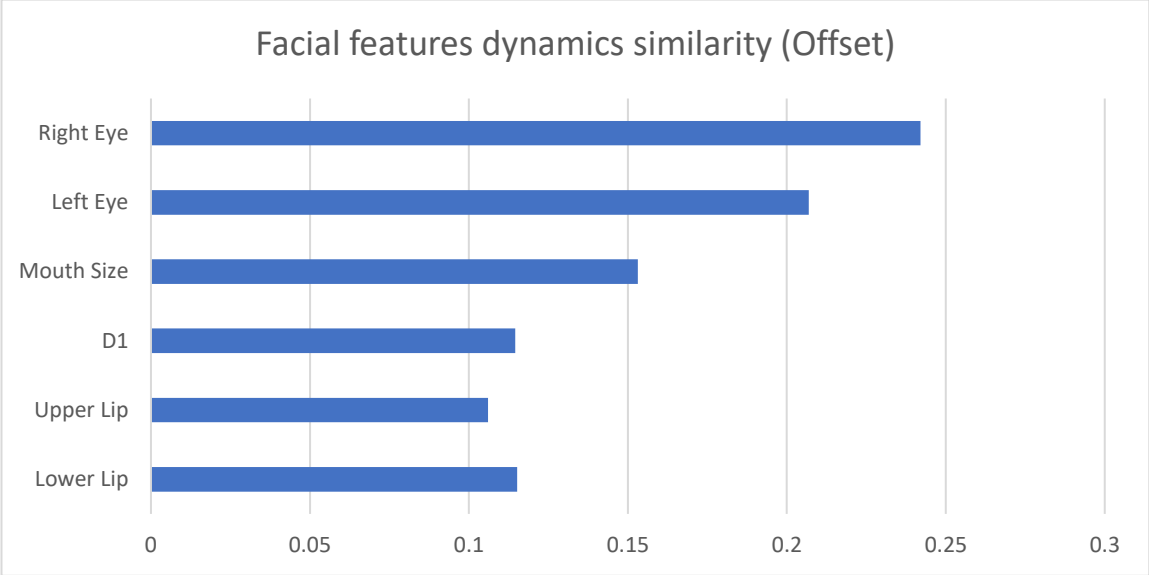


Figure 7-10: Similarity of the dynamic features at the offset interval.

In the second experiment, we compute the average flow similarity for each smile interval. Figure 7-11 shows the average similarity of the total flow in the facial features of each subject, normalised for each smile interval.

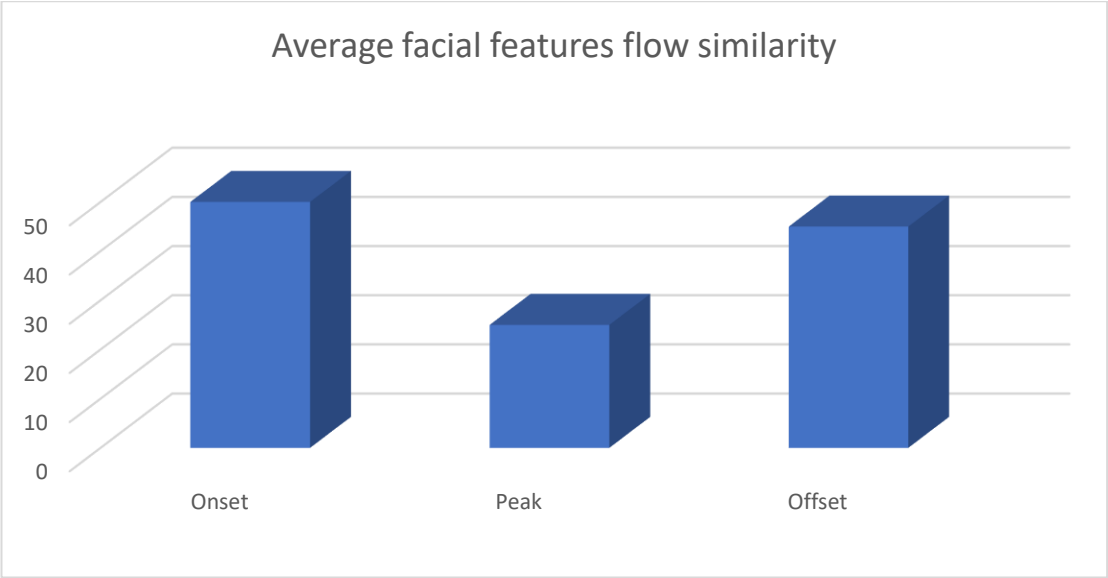


Figure 7-11: Face flow similarity within smile intervals.

Finally, we compute the time features, where we compute the similarity of time intervals between the subjects in the MUG dataset. Figure 7-12 shows the smile interval similarity. Using the total time as the signature shows that 8% of the subjects use the same timing.

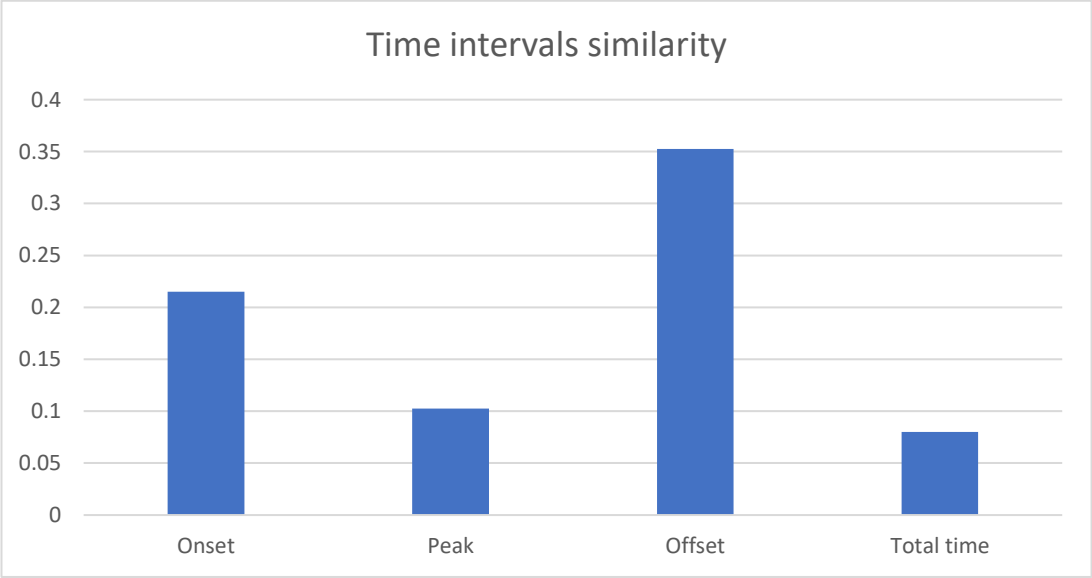


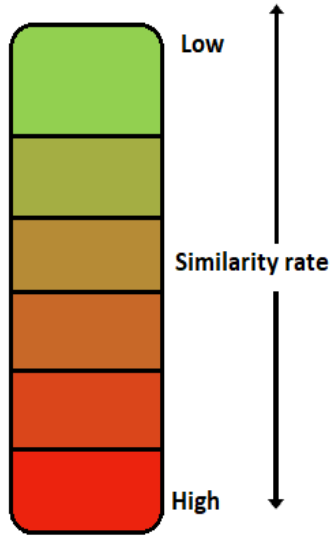
Figure 7-12: Time intervals' similarity.

As we use the flow value for computing the total displacement as shown in figure 7-11, we use it in computing order of activation. After testing the order of activation, we find a variation within the same subject in terms of the order they use facial features to formulate the smile. Additionally, the result shows that the mouth region has the highest flow value, cheeks the second and eyes the lowest. These findings divert our attention to which parts of the face have more weight in smile expression (left or right). Using the proposed method, the results show partial stability of using the left or right side of the face for the same subject. The result shows that 60% of the subjects in the MUG dataset use their left side of the face and the other 40% use the right side. Due to the weakness of this feature, we eliminated it from our computation of similarity matrix.

For validation of our approach, we compare all the feature clusters for all the subjects in the MUG dataset. We use (β, γ) to identify the overall similarity for each subject in the dataset. Table 7-2 shows a similarity heat map for the subject where the same subject source has the highest scores.

Table 7-2: Similarity heat map.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
S 1	100	11.1111	8.3333	0	0	0	0	11.1111	0	16.6667	33.3333	0	0	0	0	16.6667	0	4.16667	22.2222	16.6667
S 2	0	100	4.16667	0	0	0	0	0	0	0	0	0	33.3333	12.5	0	5.55556	0	0	5.55556	27.7778
S 3	33.3333	27.7778	62.5	33.3333	44.4444	11.1111	33.3333	33.3333	33.3333	50	50	33.3333	38.8889	25	38.8889	44.4444	25	50	38.8889	50
S 4	50	11.1111	41.6667	100	55.5556	11.1111	22.2222	66.6667	0	38.8889	33.3333	22.2222	33.3333	12.5	61.1111	44.4444	37.5	45.8333	38.8889	22.2222
S 5	83.3333	22.2222	29.1667	27.7778	75	16.6667	16.6667	61.1111	0	44.4444	27.7778	0	22.2222	25	16.6667	27.7778	37.5	50	50	27.7778
S 6	0	5.55556	0	0	11.1111	75	0	27.7778	0	11.1111	0	0	5.55556	0	0	5.55556	8.3333	12.5	11.1111	5.55556
S 7	33.3333	11.1111	12.5	22.2222	16.6667	11.1111	75	33.3333	16.6667	33.3333	33.3333	27.7778	16.6667	12.5	16.6667	27.7778	25	16.6667	16.6667	22.2222
S 8	33.3333	16.6667	12.5	16.6667	27.7778	11.1111	16.6667	75	16.6667	27.7778	27.7778	16.6667	16.6667	12.5	16.6667	16.6667	16.6667	20.8333	22.2222	22.2222
S 9	0	0	4.16667	16.6667	5.55556	0	22.2222	16.6667	100	11.1111	16.6667	22.2222	0	0	0	11.1111	12.5	0	0	0
S 10	16.6667	11.1111	20.8333	0	5.55556	11.1111	0	22.2222	0	100	11.1111	0	0	0	0	16.6667	8.3333	12.5	11.1111	44.4444
S 11	66.6667	27.7778	25	33.3333	61.1111	11.1111	33.3333	50	16.6667	61.1111	94.4444	22.2222	22.2222	25	33.3333	61.1111	12.5	37.5	55.5556	33.3333
S 12	0	0	0	0	0	0	50	0	33.3333	0	33.3333	72.2222	0	0	11.1111	5.55556	0	0	0	0
S 13	0	0	0	0	0	0	0	0	0	0	0	0	72.2222	16.6667	0	0	0	0	0	16.6667
S 14	0	0	4.16667	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0
S 15	66.6667	16.6667	8.3333	11.1111	61.1111	22.2222	38.8889	66.6667	0	66.6667	33.3333	5.55556	11.1111	4.16667	100	44.4444	37.5	41.6667	33.3333	33.3333
S 16	66.6667	16.6667	25	16.6667	33.3333	16.6667	16.6667	38.8889	0	44.4444	44.4444	16.6667	16.6667	12.5	27.7778	100	25	20.8333	22.2222	27.7778
S 17	16.6667	11.1111	37.5	55.5556	16.6667	5.55556	33.3333	33.3333	0	16.6667	16.6667	33.3333	44.4444	25	16.6667	16.6667	100	16.6667	27.7778	16.6667
S 18	16.6667	44.4444	29.1667	16.6667	66.6667	27.7778	0	55.5556	0	50	16.6667	5.55556	16.6667	25	16.6667	33.3333	29.1667	75	61.1111	72.2222
S 19	16.6667	22.2222	12.5	11.1111	27.7778	16.6667	0	0	0	16.6667	27.7778	0	0	20.8333	5.55556	33.3333	0	25	100	38.8889
S 20	25	33.3333	16.6667	27.7778	33.3333	11.1111	0	27.7778	0	33.3333	11.1111	0	33.3333	25	11.1111	27.7778	16.6667	41.6667	44.4444	100



7.4 Conclusions

In this chapter, we investigate the possibility of using the smile expression as a biometric. Using machine-learning techniques, such as PCA and CNN, we gain very high recognition rates where these techniques use a huge number of features and gain a 99% recognition rate. On the other hand, we notice that both machine-learning techniques used here make use of a lot of features which are not related to the dynamic of the smile which conflict with our main research question of this chapter, i.e. can we use smile dynamic as a biometric?

In this chapter, we have presented a novel algorithm to study if the dynamic features of the smile can be used as a biometric. Our approach is to study only the dynamic characteristics of the facial features through the smile expression. First, we identify the smile interval which represents different parts of the smile in terms of time such as the start point, peak points and the end of the smile. Using time intervals, we focus our analysis on different dynamic features which include, upper and lower lips, mouth area, eyes, the distance between the mouth corners, the flow of the face and time of smile in terms of smile interval. The proposed algorithm identifies the smile intervals showing a 95% correct detection when tested with the manually coded intervals in the MUG dataset.

For classification, we present the feature interval which is a 2D cluster that captures different variations of features through the smile expression. Feature intervals are computed using the average and the standard deviation for different smiles for the same subject and used to measure the degree of similarity between different subjects.

Our results show that humans smile expression is stable which includes the mouth size and smile length. Additionally, we show a degree of similarity among the features in the subjects in the MUG dataset which signify the uniqueness of the smile expression. Further, the results indicate that the dynamics can be used to identify different subjects uniquely. Finally, by analysing the facial feature activation we conclude that all humans have a similar order of facial features when expressing a smile.

Thus, our results also indicate that smile expression can be used as a biometric or as a soft biometric due to the high similarity rate among the subjects in the MUG dataset. For future work, we intend to extend the number of features used and try different classification techniques to minimise the similarity rate. Furthermore, we will investigate other emotions for the possibility of using them as potential biometrics.

8 Conclusions, Limitations and Future Work

In this chapter we conclude this research and discuss the some of the potential limitation of the work carried so far on this topic. We further discuss potential future direction in which this work can be taken forward.

8.1 Conclusions

Emotions play a very important role in non-verbal communication as they provide an insight into human feelings and interactions [44]. Moreover, as humans become heavily dependent on technology, using devices such as computers, phones and tablets, it would be useful if computers could synthesise human facial expressions [44].

Facial expressions represent the emotional state caused by facial muscle movement. Moreover, facial expressions can give an insight into human interactions and can provide information on the circumstances in which interaction occurs [169]. Ekman [3] identified six universal emotions (fear, sadness, disgust, anger, surprise, happiness) among different cultures. These universal emotions represent a small subset of emotions experienced by humans. Facial expression analysis is useful for a wide range of applications, including, but not limited to, facial animation [29], human computer interaction [16, 17, 169] and robots [170]. Additionally, it gives computers human characteristics in terms of interacting with and responding to humans.

In this research, we study how facial expression are encoded with the dynamics of face during facial expressions, where we have introduced novel algorithms for detecting and analysing facial expressions. These algorithms use a set of minimal features to detect different action units and facial expressions

using the flow on the face. This is undertaken by identifying a set of connected regions of interest (ROI) to cover all possible movement of the facial features and using optical flow algorithms to measure the movement. As a result, we gain 86% correct classification of the six universal emotions (happy, surprise, anger, fear, sad and disgust) and 95% accuracy in detecting the commonly used 18 action units.

Through our research, we intensively study the smile expression. The work included in detecting posed and genuine smiles, identifying gender as well as human identity itself. We have focussed on studying the smile expression due the powerful characteristics because it is:

1) one of the six universal emotions,
2) smile expression is considered one of the most sophisticated facial expressions,
3) there are more than 18 types of smiles,
4) smiles have been the focus of a lot of psychological and social behavioural research.

The research carried on smile expression can be highlights as follows.

We have introduced a statistical model to analyse facial features movement through to identify posed and genuine smiles. This done by identifying a set of regions of interest and compute the movement for each ROI. Identifying the ROI is done by using facial landmarks computed through the CHEHRA model and then using optical flow algorithm to measure the movement. Our approach has been tested on two publicly available datasets namely the CK+ and MUG. which has data representing posed and genuine smiles respectively. Our results

show that eyes contain the most information in order to distinguish between a posed and a genuine smile. Furthermore, a real smile has more facial muscle movement compared to a fake smile. Finally, a real smile has distinguishable movement distribution around the eyes compared to a fake smile.

For identifying gender, we have studied the dynamic characteristics of the facial features through the smile expressions looking for clues of gender. This idea is supported by a set of psychological studies that show female have more expressive smiles than males. We proposed a novel algorithm for detecting gender based on measuring the spatial, area, geometric and intrinsic features which represent the dynamics of the smile. Using our approach, we successfully classify gender by looking only at the dynamic characteristics of the smile where we gain 86% in MUG dataset and 78% in the CK+ dataset classified using the KNN algorithm.

Finally, we take the concept of identifying gender using the smile dynamic a further step to identify the possibility of using smile expression as a biometric. In this research, we have studied the possibility of using the dynamic characteristics of smile expression to identify human identity. This done by measuring a set spatial, area, geometric, intrinsic and timing features in a specific time interval (smile interval). Our results indicate that the dynamic characteristics of the smile can be used as a biometric or as a soft-biometric and it shows that humans have a stable factor when expressing the smile over time. This research can be used as a foundation for creating a sophisticated biometric that we believe cannot be manipulated or duplicated.

8.2 General Limitations

There are a number of limitations of this present work which we summarise below.

A. Head movement

As mentioned in the previous chapters, identifying and correcting for head movements is a very challenging problem where it affects the detection, analysis and the classification process. As our methods are computationally driven, we consider head movement one of the main challenges.

B. Face-pose

All our analysis and the proposed method were carried on datasets containing subjects with frontal view. Thus, subjects with non-frontal views will not work properly with proposed methods as it affects the face and landmark detection algorithms as well as the classification process.

C. Environmental condition

Our analysis was performed on datasets where subjects were recorded in carefully controlled environments. Running the algorithms on data from uncontrolled environment is still a challenging problem. Other researchers working on these problems have identified these problems such face detection, landmark detection algorithms and motion detection techniques which remain challenging when subjected to difficult environmental conditions.

D. Datasets

Our analysis and detection were carried out on two datasets, namely the CK+ and the MUG datasets, which were designed and created in a controlled environment. We think if different datasets were obtained (e.g. more genuine smiles) there could be better results for discriminating between genuine and posed smiles. Moreover, for facial biometrics, using the smile dynamics, there are no publicly validated datasets and hence the accuracy of our results in this regard remains to be challenged.

8.3 Future Work

In this section we discuss in detail the future work for each chapter as follows.

In chapter 4, we show how facial expressions are encoded within the dynamics of facial expressions, particularly in the smile. The proposed method can be enhanced further by adding both geometric and appearance information to overcome the limitation mentioned in Section 4.4. For example, using the active shape model (ASM) for identifying the geometric features and the active appearance model (AAM) for analysing the texture of the facial expressions may provide improvements. Similarly, classification can be improved using algorithms such as KNN, SVM and other similar classification algorithms. Additionally, the results can be extended to various other face poses apart from the frontal face. Finally, more powerful machine learning algorithms such as the use of CNN can be proposed. This can be used with large datasets to train the neural network

and to compare the hybrid model described above with the output of the CNN in terms of accuracy and performance.

In chapter 5, we presented a statistical model used to analyse posed and genuine smiles. In our future work, we can use the findings of this research as a ground rule for creating an application for detecting fake and real smiles in real life environments where we propose a texture analysis algorithm to be performed on the eyes area which identifies the wrinkles and the movement of eyes more accurately. A Gabor function based approach, for example, can be utilised to make such enhancements. Other methods such as the active appearance model and wrinkle detection techniques will equally suit too. Additionally, we intend to study other emotions and their relationship with facial features arising from such emotions to quantify posed and genuine smiles.

In chapter 6, we have shown that gender is encoded within the dynamic characteristics of the smile expressions. For future work, we intend to extend the number of features used which include more geometric distances and facial features shape representation in order to increase the feature space. This includes adding the active shape model to the system and comparing it with the proposed features to represent the dynamic features more accurately. For classification, we can use the CNN and compare it with other classification algorithms. Finally, we intend to study other emotions for gender difference. It is noteworthy to out point out that psychological studies have shown that other facial expressions such as fear and anger are considered to be more masculine and hence there is an avenue to undertake computer based on such facial expressions.

In chapter 7, we have shown the possibility of using the dynamic characteristics of the smile expression as a potential biometric. Future work includes can go into four possible directions. First, construing a new dataset which contains unique human identities such as the facial expressions from identical twins expressing emotions, as there are no datasets such publicly available datasets. Second, it is plausible to design a real-time application using the proposed method and test the approach on real data by taking into consideration this as soft biometric. Third, to find a way to use deep machine learning algorithms such as CNN and DCNN to identify the dynamic characteristics of facial expressions could be interesting. To achieve this, we suggest removing texture data from the face images to account for possible noise in the images which can then utilised to extract the dynamic features to train the CNN. Finally, once the above step is successful for the smile expression, we can investigate the use of other facial emotions such as fear, anger and surprise to look for biometric clues in them.

References

[1] B. Fehr, and J. A. Russell, “Concept of emotion viewed from a prototype perspective,” *Journal of Experimental Psychology,* vol. 113, no. 3, pp. 464-486, 1984.

[2] M. Grimm, D. Dastidar, and K. Kroschel, "Recognizing emotions in spontaneous facial expressions," in *Proceedigns of the International Conference on Intelligent Systems and Computing (ISYC)*, pp. 865–868, 2006.

[3] P. Ekman, and W. V. Friesen, “Constants across cultures in the face and emotion,” *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124-129, 1971.

[4] J. LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*: Simon And Schuster, 1998.

[5] W. B. Cannon, *Bodily Changes in Pain, Hunger, Fear, and Rage: An Account of Recent Researches into the Function of Emotional Excitement*, D. Appleton, 1916.

[6] humanillnesses.com, "Emotions," 25/10, 2015. http://www.humanillnesses.com/Behavioral-Health-Br-Fe/Emotions.html.

[7] W. James, "*Principles of Psychology*," *Principles of Psychology*, 1890.

[8] H. K. Shergill, "*Psychology, part 1*," PHI Learning, pp. 314-317, 2010.

[9] S. Schachter, and J. Singer, “Cognitive, social, and physiological determinants of emotional state,” *Psychological Review*, vol. 69, no. 5, pp. 379-399, 1962.

[10] F. Metze, A. Batliner, F. Eyben, T. Polzehl, B. Schuller, and S. Steidl, "Emotion recognition using imperfect speech recognition," in *Eleventh Annual Conference of the International Speech Communication Association*, pp. 478-481, 2010.

[11] K. Schindler, L. Van Gool, and B. de Gelder, “Recognizing emotions expressed by body pose: A biologically inspired neural model.” *Neural Networks*, vol. 21, no. 9, pp. 1238-1246, 2008.

[12] M. M. Parker, "*Understanding Psychology*," Redding: Horizon Textbook Publishing, pp. 327 -328, 2007.

[13] W. G. Parrott, "*Emotions in Social Psychology," Essential Readings*: Psychology Press, 2001.

[14] A. Ortony, and T. J. Turner, “What's basic about basic emotions?,” *Psychological Review*, vol. 97, no. 3, pp. 315-331, 1990.

[15] B. Jahne, *Computer Vision And Applications: A Guide For Students and Practitioners*: Academic Press, 2000.

[16] R. B. Fisher, T. P. Breckon, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco, and C. K. Williams, *Dictionary of Computer Vision and Image Processing*: John Wiley & Sons, 2013.

[17] A. Jaimes, and N. Sebe, “Multimodal human–computer interaction: A survey,” *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 116-134, 2007.

[18] D. M. Gavrila, “The visual analysis of human movement: a survey,” *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82-98, 1999.

[19] C. G. Wolf, “Can people use gesture commands?,” *ACM SIGCHI Bulletin*, vol. 18, no. 2, pp. 73-74, 1986.

[20] J. L. Raheja, R. Shyam, U. Kumar, and P. B. Prasad, "Real-time robotic hand control using hand gestures," in *Second IEEE International Conference on Machine Learning and Computing (ICMLC)*, pp. 12-16, 2010.

[21] A. R. Naghsh-Nilchi, and M. Roshanzamir, “An efficient algorithm for motion detection based facial expression recognition using optical flow,” in *Proceedings of the World Academy of Science*, *Engineering and Technology*, vol. 20, pp. 23-28, 2006.

[22] A. Haro, M. Flickner, and I. Essa, "Detecting and tracking eyes by using their physiological properties, dynamics, and appearance," in *the IEEE coference on Computer Vision and Pattern Recognition*, vol. 1, pp. 163-168, 2000.

[23] F. Timm, and E. Barth, "Accurate eye centre localisation by means of gradients," in *Proceedigs of the Sixth International Conference in Computer Vision Theory and Applications*, pp. 125-130, 2011.

[24] P. Viola, and M. J. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[25] A. Metallinou, S. Lee, and S. Narayanan, "Audio-visual emotion recognition using gaussian mixture models for face and voice," in *Tenth IEEE International Symposium Multimedia (ISM 2008)*, pp. 250-257, 2008.

[26] B. Azar, “What’s in a face,” *Monitor on Psychology*, vol. 31, no. 1, pp. 44-45, 2000.

[27] W. E. Rinn, “The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions,” *Psychological Bulletin*, vol. 95, no. 1, pp. 52-72, 1984.

[28] P. Ekman, and W. V. Friesen, *Manual for the Facial Action Coding System*: Consulting Psychologists Press, 1978.

[29] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion recognition in human-computer interaction,” *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, 2001.

[30] C.-H. Hjortsjö, *Man's face and mimic language*: Studen litteratur, 1969.

[31] P. Ekman, W. Irwin, and E. Rosenberg, “The emotional facial action coding system *(EMFACS),” Unpublished manuscript, University of California at San Francisco*, 1994.

[32] A. Freitas-Magalhães, “Microexpression and macroexpression,” *Encyclopedia of Human Behavior*, vol. 2, pp. 173-178, 2012.

[33] W. V. Friesen, and P. Ekman, “EMFACS-7: Emotional facial action coding system,” *Unpublished manuscript, University of California at San Francisco*, vol. 2, 1983.

[34] P. Ekman, and E. L. Rosenberg, *What The Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*: Oxford University Press, 1997.

[35] M. Pardàs, and A. Bonafonte, “Facial animation parameters extraction and expression recognition using hidden Markov models,” *Signal Processing, Image Communication*, vol. 17, no. 9, pp. 675-688, 2002.

[36] E. Petajan, "MPEG-4 face and body animation coding applied to HCI," *Real-Time Vision for Human-Computer Interaction*, Springer, pp. 249-268, 2005.

[37] W. Mattheyses, and W. Verhelst, “Audiovisual speech synthesis: an overview of the state-of-the-art,” *Speech Communication*, vol. 66, pp. 182-217, 2015.

[38] L. Zalewski, and S. Gong, "2d statistical models of facial expressions for realistic 3d avatar animation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR 2005),* vol. 2, pp. 217-222, 2005.

[39] A. Wojdeł, and L. J. Rothkrantz, "Parametric generation of facial expressions based on FACS," *Computer Graphics Forum*, vol. 24, no. 4, pp. 743-757, 2005.

[40] A. Al-dahoud, and H. Ugail, "A method for location based search for enhancing facial feature detection," *Advances in Computational Intelligence Systems*, Springer, pp. 421-432, 2017.

[41] H. Ugail, and A. Al-dahoud, “Is gender encoded in the smile? A computational framework for the analysis of the smile driven dynamic face for gender recognition,” *The Visual Computer*, vol. 34, no. 9, pp. 1243–1254, 2018.

[42] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 94-101, 2010.

[43] N. Aifanti, C. Papachristou, and A. Delopoulos, "The MUG facial expression database," in *IEEE 11th International Workshop on Image Analysis For Multimedia Interactive Services (WIAMIS)*, pp. 1-4, 2010.

[44] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, “Static and dynamic 3D facial expression recognition: a comprehensive survey,” *Image and Vision Computing*, vol. 30, no. 10, pp. 683-697, 2012.

[45] D. Datcu, and L. J. Rothkrantz, “The use of active appearance model for facial expression recognition in crisis environments,” in *Proceedigns of the ISCRAM 2007*, pp. 515-524, 2007.

[46] B. Fasel, and J. Luettin, “Automatic facial expression analysis: a survey,” *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[47] P. Dulguerov, F. Marchal, D. Wang, and C. Gysin, “Review of objective topographic facial nerve evaluation methods,” *American Journal of Otology*, vol. 20, no. 5, pp. 672-678, 1999.

[48] R. Koenen, *Mpeg-4 Project Overview. International Organisation for Standartistion*, ISO/IECJTC1/SC29/WG11, La Baule, Octorber, 2000.

[49] F. Daniyal, P. Nair, and A. Cavallaro, "Compact signatures for 3d face recognition under varying expressions," in *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'09)*, pp. 302-307, 2009.

[50] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer, “Meta-analysis of the first facial expression recognition challenge,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 42, no. 4, pp. 966-979, 2012.

[51] C. Ravat, and S. Solanki, “Facial Expression Recognition using Convolutional Neural Networks,” *International Journal of Scientific Research in Science*, vol. 4, no. 4, pp. 1486-1489, 2018.

[52] Z. Zhang, P. Luo, C. C. Loy, and X. J. I. J. o. C. V. Tang, “From facial expression recognition to interpersonal relation prediction,” *International Journal of Computer Vision*, vol. 126, no. 5, pp. 550-569, 2018.

[53] K. Zhang, Y. Huang, Y. Du, and L. J. I. T. o. I. P. Wang, “Facial expression recognition based on deep evolutional spatial-temporal networks,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4193-4203, 2017.

[54] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. J. P. R. Oliveira-Santos, “Facial expression recognition with convolutional neural networks: coping with few data and the training sample order,” *Pattern Recognition*, vol. 61, pp. 610-628, 2017.

[55] H. Ding, S. K. Zhou, and R. Chellappa, "Facenet2expnet: regularizing a deep face recognition net for expression recognition," in *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017),* pp. 118-126, 2017.

[56] P. Rodriguez, G. Cucurull, J. Gonzalez, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, and F. X. J. I. t. o. c. Roca, “Deep pain: exploiting long short-term memory networks for facial expression classification,” *IEEE Transactions On Cybernetics*, no. 99, pp. 1-11, 2017.

[57] A. Besinger, T. Sztynda, S. Lal, C. Duthoit, J. Agbinya, B. Jap, D. Eager, and G. Dissanayake, “Optical flow based analyses to detect emotion from human facial image data,” *Expert Systems with Applications*, vol. 37, no. 12, pp. 8897-8902, 2010.

[58] B. D. Lucas, and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedigns of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[59] S. Koelstra, and M. Pantic, "Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics," in *8th IEEE International Conference on Automatic Face & Gesture Recognition (FG'08),* pp. 1-8, 2008.

[60] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, “Automatic recognition of facial actions in spontaneous expressions,” *Journal of Multimedia*, vol. 1, no. 6, pp. 22-35, 2006.

[61] S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. De la Torre, and J. Cohn, "AAM derived face representations for robust facial action recognition," in *7th IEEE International Conference Automatic Face and Gesture Recognition (FGR 2006)*, pp. 155-160, 2006.

[62] Y.-I. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, 2001.

[63] M. F. Valstar, and M. Pantic, "Combined support vector machines and hidden markov models for modeling facial action temporal dynamics," in *International Workshop on Human-Computer Interaction*, Springer, Berlin, Heidelberg, pp. 118-127, 2007.

[64] Y. Yacoob, and L. S. Davis, “Recognizing human facial expressions from long image sequences using optical flow,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636-642, 1996.

[65] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, “Classifying facial actions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.

[66] J. N. Bassili, “Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face,” *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2049- 2058, 1979.

[67] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 46-53, 2000.

[68] M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, “Automatic recognition of facial actions in spontaneous expressions,” *Journal of Multimedia*, vol. 1, no. 6, pp. 22-35, 2006.

[69] Z. Zhang, “Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, no. 6, pp. 893-911, 1999.

[70] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, “Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders,” *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 237-256, 2011.

[71] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.

[72] M. Perron, and A. Roy-Charland, “Analysis of eye movements in the judgment of enjoyment and non-enjoyment smiles,” *Frontiers in Psychology*, vol. 4, pp. 659-670, 2013.

[73] Gunnery, Sarah D., and Mollie A. Ruben. "Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis," *Cognition and Emotion,* vol. 30, no. 3, pp. 501-515, 2016.

[74] P. Ekman, R. J. Davidson, and W. V. Friesen, “The Duchenne smile: emotional expression and brain physiology: II,” *Journal of Personality and Social Psychology*, vol. 58, no. 2, pp. 342-353, 1990.

[75] X. Mai, Y. Ge, L. Tao, H. Tang, C. Liu, and Y.-J. Luo, “Eyes are windows to the Chinese soul: evidence from the detection of real and fake smiles,” *PloS One*, vol. 6, no. 5, pp. 1-6, 2011.

[76] U. Dimberg, and L.-O. Lundquist, “Gender differences in facial reactions to facial expressions,” *Biological Psychology*, vol. 30, no. 2, pp. 151-159, 1990.

[77] U. Dimberg, “Facial electromyographic reactions and autonomic activity to auditory stimuli,” *Biological Psychology*, vol. 31, no. 2, pp. 137-147, 1990.

[78] V. Surakka, and J. K. Hietanen, “Facial and emotional reactions to Duchenne and non-Duchenne smiles,” *International Journal of Psychophysiology*, vol. 29, no. 1, pp. 23-33, 1998.

[79] A. Van Boxtel, "Facial EMG as a tool for inferring affective states," in *Proceedings of the Measuring Behavior*, pp. 104-108, 2010.

[80] U. Dimberg, M. Thunberg, and K. Elmehed, “Unconscious facial reactions to emotional facial expressions,” *Psychological Science*, vol. 11, no. 1, pp. 86-89, 2000.

[81] C. Sittel, and E. Stennert, “Prognostic value of electromyography in acute peripheral facial nerve palsy,” *Otology & Neurotology*, vol. 22, no. 1, pp. 100-104, 2001.

[82] M. J. Bernstein, D. F. Sacco, C. M. Brown, S. G. Young, and H. M. Claypool, “A preference for genuine smiles following social exclusion,” *Journal of Experimental Social Psychology*, vol. 46, no. 1, pp. 196-199, 2010.

[83] R. Soussignan, and B. Schaal, “Forms and social signal value of smiles associated with pleasant and unpleasant sensory experience,” *Ethology*, vol. 102, no. 8, pp. 1020-1041, 1996.

[84] M. G. Calvo, A. Gutiérrez-García, P. Avero, and D. Lundqvist, “Attentional mechanisms in judging genuine and fake smiles: eye-movement patterns,” *Emotion*, vol. 13, no. 4, pp. 792-802, 2013.

[85] V. Manera, M. Del Giudice, E. Grandi, and L. Colle, “Individual differences in the recognition of enjoyment smiles: no role for perceptual–attentional factors and autistic-like traits,” *Frontiers in Psychology*, vol. 2, pp. 143-152, 2011.

[86] P. Gosselin, M. Beaupré, and A. Boissonneault, “Perception of genuine and masking smiles in children and adults: Sensitivity to traces of anger,” *The Journal of Genetic Psychology*, vol. 163, no. 1, pp. 58-71, 2002.

[87] P. Wu, W. Wang, and H. Liu, "Methods of recognizing true and fake smiles by using AU6 and AU12 in a holistic way," in *Proceedings of the 2013 Chinese Intelligent Automation Conference*, Springer, Berlin, Heidelberg, pp. 603-613, 2013.

[88] M. Nakano, Y. Mitsukura, M. Fukumi, and N. Akamatsu, "True smile recognition system using neural networks," in *Proceedings of the 9th IEEE International Conference on Neural Information Processing(ICONIP'02)*, vol. 2, pp. 650-654, 2002.

[89] E. L. Abel, and M. L. Kruger, “Smile intensity in photographs predicts longevity,” *Psychological Science*, vol. 21, no. 4, pp. 542-544, 2010.

[90] P. Bilinski, A. Dantcheva, and F. Brémond, "Can a smile reveal your gender?," in *IEEE International Conference of Biometrics Special Interest Group (BIOSIG),* pp. 1-6, 2016.

[91] D. Matsumoto, and B. Willingham, “Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals,” *Journal of Personality and Social Psychology*, vol. 96, no. 1, pp. 1-10, 2009.

[92] F. M. Deutsch, D. LeBaron, and M. M. Fryer, “What is in a smile?,” *Psychology of Women Quarterly*, vol. 11, no. 3, pp. 341-352, 1987.

[93] U. Hess, R. B. Adams Jr, and R. E. Kleck, “Facial appearance, gender, and emotion expression,” *Emotion*, vol. 4, no. 4, pp. 378-388, 2004.

[94] M.-F. Liébart, C. Fouque-Deruelle, A. Santini, F.-L. Dillier, V. Monnet-Corti, J.-M. Glise, and A. Borghetti, “Smile line and periodontium visibility,” *Periodontal Practice Today*, vol. 1, no. 1, pp. 17-25, 2004.

[95] R. W. Simon, and L. E. Nath, “Gender and emotion in the United States: do men and women differ in self-reports of feelings and expressive behavior?,” *American Journal of Sociology*, vol. 109, no. 5, pp. 1137-1176, 2004.

[96] I. J. Livingston, L. E. Nacke, and R. L. Mandryk, "Influencing experience: the effects of reading game reviews on player experience," in *International Conference on Entertainment Computing*, Springer, pp. 89-100, 2011.

[97] E. Cashdan, “Smiles, speech, and body posture: how women and men display sociometric status and power,” *Journal of Nonverbal Behavior*, vol. 22, no. 4, pp. 209-228, 1998.

[98] N. J. Briton, and J. A. Hall, “Gender-based expectancies and observer judgments of smiling,” *Journal of Nonverbal Behavior*, vol. 19, no. 1, pp. 49-65, 1995.

[99] S. Kalam, and G. Guttikonda, “Gender classification using geometric facial features,” *International Journal of Computer Applications*, vol. 85, no. 7, pp. 32-37, 2014.

[100] D. P. Lale, and K. J. Karande, “Gender classification using facial features,” in *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 5, no. 9, pp. 2227-2231, 2016.

[101] A. Dantcheva, and F. Brémond, “Gender estimation based on smile-dynamics,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 3, pp. 719-729, 2017.

[102] P. Rai, and P. Khanna, "Appearance based gender classification with PCA and (2D) 2 PC A on approximation face image," in *9th IEEE International Conference Industrial and Information Systems (ICIIS)*, pp. 1-6, 2014.

[103] H.-C. Lian, and B.-L. Lu, “Multi-view gender classification using local binary patterns and support vector machines,” *Advances in Neural Networks-ISNN 2006*, pp. 202-209, 2006.

[104] S. Mozaffari, H. Behravan, and R. Akbari, "Gender classification using single frontal image per person: combination of appearance and geometric based features," in *20th IEEE International Conference Pattern Recognition (ICPR)*, pp. 1192-1195, 2010.

[105] L. Lu, and P. Shi, "A novel fusion-based method for expression-invariant gender classification," in *IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP 2009)*, pp. 1065-1068, 2009.

[106] Z. Xu, L. Lu, and P. Shi, "A hybrid approach to gender classification from face images," in *9th IEEE International Conference on Pattern Recognition (ICPR 2008)*, pp. 1-4, 2008.

[107] T. Jabid, M. H. Kabir, and O. Chae, "Gender classification using local directional pattern (LDP)," in *IEEE 20th International Conference on Pattern Recognition (ICPR)*, pp. 2162-2165, 2010.

[108] R. Jafri, and H. R. Arabnia, “A survey of face recognition techniques,” *Jips*, vol. 5, no. 2, pp. 41-68, 2009.

[109] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, “Face recognition from a single image per person: a survey,” *Pattern Recognition*, vol. 39, no. 9, pp. 1725-1745, 2006.

[110] T. Kanade, “Picture processing system by computer complex and recognition of human faces,” *PhD Thesis, Kyoto University, Department of Informatics and Science*, 1974.

[111] R. Brunelli, and T. Poggio, “Face recognition: Features versus templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[112] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-based face recognition using mixture-distance," in *IEEE Computer Society Conference Computer Vision and Pattern Recognition*, *Proceedings CVPR'96,* pp. 209-216, 1996.

[113] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.

[114] G. Sukthankar, *Face recognition: a critical look at biologically-inspired approaches*: Carnegie Mellon University, The Robotics Institute, 2000.

[115] L. C. Paul, and A. Al Sumam, “Face recognition using principal component analysis method,” *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, no. 9, pp. 135-139, 2012.

[116] N. Kar, M. K. Debbarma, A. Saha, and D. R. Pal, “Study of implementing automated attendance system using face recognition technique,” *International Journal of Computer and Communication Engineering*, vol. 1, no. 2, pp. 100-104, 2012.

[117] V. Maheshkar, S. Agarwal, V. K. Srivastava, and S. Maheshkar, “Face recognition using geometric measurements, directional edges and directional multiresolution information,” *Procedia Technology*, vol. 6, pp. 939-946, 2012.

[118] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.

[119] M. A. Kashem, M. N. Akhter, S. Ahmed, and M. M. Alam, “Face recognition system based on principal component analysis (PCA) with back propagation neural networks (BPNN),” *Canadian Journal on Image Processing and Computer Vision*, vol. 2, no. 4, pp. 36-45, 2011.

[120] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV 2005)*, vol. 1, pp. 786-791, 2005.

[121] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, "Face recognition based on geometric features using Support Vector Machines," in *6th IEEE International on Conference Soft Computing and Pattern Recognition (SoCPaR)*, pp. 89-95, 2014.

[122] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, "Combined local features selection for face recognition based on Naïve Bayesian classification," in 1*3th IEEE International Conference on Hybrid Intelligent Systems (HIS)*, pp. 240-245, 2013.

[123] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, “Automatic gait recognition by symmetry analysis,” *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2175-2183, 2003.

[124] S. Tulyakov, T. Slowe, Z. Zhang, and V. Govindaraju, "Facial expression biometrics using tracker displacement features," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1-5, 2007.

[125] J. F. Cohn, K. Schmidt, R. Gross, and P. Ekman, "Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification," in *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, IEEE Computer Society, pp. 491-497, 2002.

[126] K. L. Schmidt, and J. F. Cohn, "Dynamics of facial expression: Normative characteristics and individual differences," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 547-550, 2001.

[127] S. Zafeiriou, C. Zhang, and Z. Zhang, “A survey on face detection in the wild: past, present and future,” *Computer Vision and Image Understanding*, vol. 138, pp. 1-24, 2015.

[128] Z. Kalal, K. Mikolajczyk, and J. Matas, "Face-tld: tracking-learning-detection applied to faces," in *17th IEEE International Conference on Image Processing (ICIP)*, pp. 3789-3792, 2010.

[129] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: a literature survey,” *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399-458, 2003.

[130] E. Hjelmås, and B. K. Low, “Face detection: a survey,” *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236-274, 2001.

[131] P. I. Wilson, and J. Fernandez, “Facial feature detection using Haar classifiers,” *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127-133, 2006.

[132] S. K. Singh, D. Chauhan, M. Vatsa, and R. Singh, “A robust skin color based face detection algorithm,” *Tamkang Journal of Science and Engineering*, vol. 6, no. 4, pp. 227-234, 2003.

[133] Y.-T. Pai, S.-J. Ruan, M.-C. Shie, and Y.-C. Liu, "A simple and accurate color face detection algorithm in complex background," in *IEEE International Conference on Multimedia and Expo,* pp. 1545-1548, 2006.

[134] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, 1998.

[135] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *IEEE confernce on Computer Vision and Pattern Recognition*, pp. 130-136, 1997.

[136] S. T. Y. Ping, C. H. Weng, and B. Lau, “Face detection through template matching and color segmentation,” *EE 368 Final Project, Stanford University*, 2003.

[137] S. Milborrow, and F. Nicolls, "Locating facial features with an extended active shape model," *Computer Vision–ECCV 2008*, Springer, pp. 504-513, 2008.

[138] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 681-685, 2001.

[139] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Comparing Active Shape Models with Active Appearance Models," in *BMVC*, vol. 99, no. 1, pp. 173-182, 1999.

[140] X. Chen, and W. Cheng, “Facial expression recognition based on edge detection,” *International Journal of Computer Science and Engineering Survey*, vol. 6, no. 2, pp. 1-9, 2015.

[141] S. Srivastava, “Real time facial expression recognition using a novel method,” *International Journal of Multimedia & Its Applications (IJMA)*, vol. 4, no. 2, pp. 49-57, 2012.

[142] J. So, S. Han, H.-C. Shin, and Y. Han, “Active appearance model-aased facial expression recognition using face symmetrical property in occluded face image,” *International Journal of Advances in Mechanical and Automobile Engineering*, vol. 1, no. 1, pp. 92-95, 2014.

[143] K. Hogan, *Can't get through: eight barriers to communication*: Pelican Publishing, 2003.

[144] Y.-Q. J. I. P. O. L. Wang, “An analysis of the Viola-Jones face detection algorithm,” *Image Processing Online*, vol. 4, pp. 128-148, 2014.

[145] J. F. Rüdiger, and P. V. Boesen, "Measurement of Facial Muscle EMG Potentials for Predictive Analysis Using a Smart Wearable System and Method," US Patent Application 15/703, 811, 2018.

[146] J. Perdiz, G. Pires, and U. J. Nunes, "Emotional state detection based on EMG and EOG biosignals: A short survey," in *IEEE 5th Portuguese Meeting Bioengineering (ENBENG)*, pp. 1-4, 2017.

[147] G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Scandinavian Conference on Image Analysis*, Springer, Berlin, Heidelberg, pp. 363-370, 2003.

[148] P. D. Bryn Farnsworth. "Facial Action Coding System (FACS) – A Visual Guidebook," 1/1, 2017; https://imotions.com/blog/facial-action-coding-system/.

[149] "About OpenCv," 20/11, 2015; http://opencv.org/about.html.

[150] M. W. Kraus, and T.-W. D. Chen, “A winning smile? Smile intensity, physical dominance, and fighter performance,” *Emotion*, vol. 13, no. 2, pp. 270-279, 2013.

[151] M. J. Hertenstein, C. A. Hansel, A. M. Butts, and S. N. Hile, “Smile intensity in photographs predicts divorce later in life,” *Motivation and Emotion*, vol. 33, no. 2, pp. 99-105, 2009.

[152] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*: W. W. Norton, 2009.

[153] H. Dibeklioğlu, A. A. Salah, and T. Gevers, "Are you really smiling at me? Spontaneous versus posed enjoyment smiles," in *European Conference on Computer Vision*, Springer, Berlin, Heidelberg, pp. 525-538, 2012.

[154] G. Littlewort-Ford, M. S. Bartlett, and J. R. Movellan, “Are your eyes smiling? Detetcting genuine smiles with support vector machines and Gabor wavelet,” in *Proceedings of the 8th Joint Symposium on Neural Computation*, pp. 1-9, 2001.

[155] P. Thibault, M. Levesque, P. Gosselin, and U. J. S. P. Hess, “The Duchenne marker is not a universal signal of smile authenticity–but it can be learned!,” *Social Psychology*, vol. 43, pp. 215-221, 2012.

[156] W. M. Brown, and C. Moore, "Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary analysis," *Advances in Psychological Research,* vol. 11. pp. 59–78, 2002.

[157] P. Thibault, P. Gosselin, M.-L. Brunel, and U. Hess, “Children’s and adolescents’ perception of the authenticity of smiles,” *Journal of Experimental Child Psychology*, vol. 102, no. 3, pp. 360-367, 2009.

[158] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1859-1866, 2014.

[159] M. G. Frank, P. Ekman, and W. V. Friesen, “Behavioral markers and recognizability of the smile of enjoyment,” *Journal of Personality and Social Psychology*, vol. 64, no. 1, pp. 83-93, 1993.

[160] Z. Živković, "Improved adaptive Gausian mixture model for background subtraction," in *Proceedings of the 17th IEEE International Conference on Pattern Recognition, (ICPR 2004),* vol. 2, pp. 28-31, 2004.

[161] G. Bradski, and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*: O'Reilly Media, Inc, 2008.

[162] L. R. Rubin, “The anatomy of a smile: its importance in the treatment of facial paralysis,” *Plastic and Reconstructive Surgery*, vol. 53, no. 4, pp. 384-387, 1974.

[163] L. R. Brody, and J. A. Hall, “Gender, emotion, and expression,” *Handbook of emotions*, vol. 2, pp. 338-349, 2000.

[164] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, “Multi-modal approach for affective computing,” in *IEEE 40th International Confernce on Engineering in Medicine and Biology (EMBC)*, arXiv preprint, arXiv:1804.09452v2, 2018.

[165] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision (WACV),* pp. 1-10, 2016.

[166] G. G. Chrysos, E. Antonakos, P. Snape, A. Asthana, and S. J. I. J. o. C. V. Zafeiriou, “A comprehensive performance evaluation of deformable face tracking “in-the-wild”,” *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 198-232, 2018.

[167] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *The American Statistician*, vol. 46, no. 3, pp. 175-185, 1992.

[168] C.-F. Tsai, and C.-Y. J. P. r. Lin, “A triangle area based nearest neighbors approach to intrusion detection,” *Pattern Recognition*, vol. 43, no. 1, pp. 222-229, 2010.

[169] F. Karray, M. Alemzadeh, J. A. Saleh, and M. N. Arab, “Human-computer interaction: overview on state of the art,” *International Journal on Smart Sensing and Intelligent Systems*, vol. 1, no. 1, pp. 137-159, 2008.

[170] Y. Yang, S. S. Ge, T. H. Lee, and C. Wang, “Facial expression recognition and tracking for intelligent human-robot interaction,” *Intelligent Service Robotics*, vol. 1, no. 2, pp. 143-157, 2008.

[171] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, “Subject independent facial expression recognition with robust face detection using a convolutional neural network,” *Neural Networks*, vol. 16, no. 5-6, pp. 555-559, 2003.

[172] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[173] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, arXiv preprint, arXiv:1409.1556v6, 2015.

Appendix A: Location of ROI

Allocating ROI is carried out by applying two main steps namely, identifying the location of ROI and applying the ROI on the face image.

1) **Identifying the ROI location**

To allocate region of interest (ROI), we need to understand the facial muscles structure. Faces comprise of muscles and nerves. Generally facial movement occurs when a set of muscles and nerves are triggered resulting in facial expressions. From an anatomical point view, we can allocate different facial muscles associated with each AU. Figure A-1 (a) shows the anatomical view of facial muscle and their location in the face. Using the location of each correlated muscle location with AUs, we placed 22 ROI as shown in Figure A-1(b).

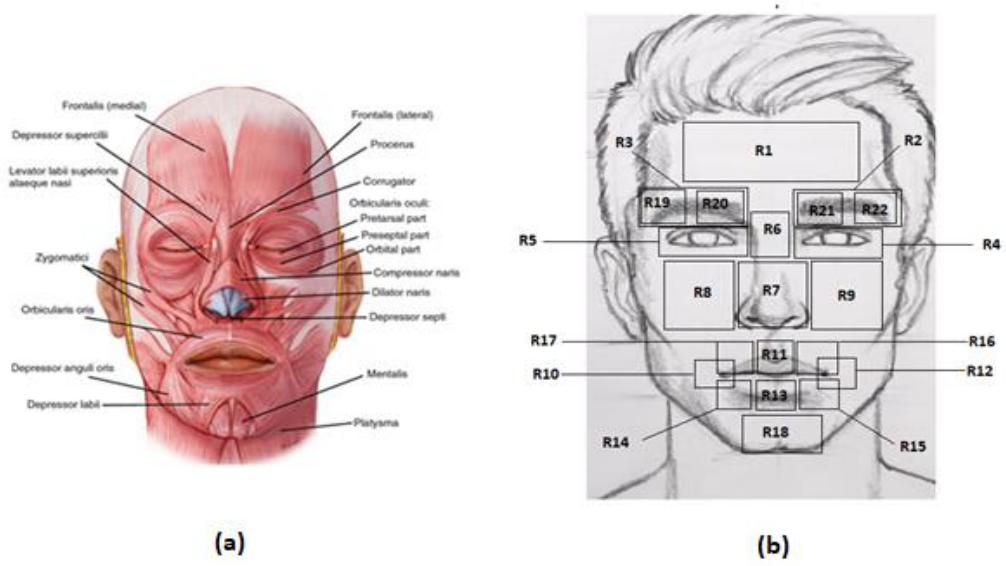


Figure A-1: (a) Anatomical structure of the face, (b) Region of the interests.

2) **Applying ROI on the Face**

Allocating the ROI is carried out by applying a set of rectangular and square areas at the approximate location of the resized face with fixed width and height. The allocation of each ROI done by studying location of the face muscles and their relationship with the corresponding AUs. Furthermore, using experiments to allocate the appropriate location and size of each ROI is done by trying different size and location to get to the best fit approximation of each ROI on CK+ dataset. Table A-1 below shows the location and size for 22 ROI.

Table A-1: ROI specification.

ROI	Start location (X, Y)		Window Size	Area Name
R 1	45	20	150 * 30	Front head
R 2	110	45	50 * 25	Right eyebrows
R 3	35	45	50 * 25	Left eyebrows
R 4	120	60	30 * 30	Right eye
R 5	50	60	30 *30	Left eye
R 6	70	90	60 * 40	Upper nose area
R 7	90	60	25 *70	Lower nose area
R 8	40	100	40 * 50	Left cheek
R 9	140	100	40 * 50	Right cheek
R 10	50	150	30 * 30	Left corner mouth
R 11	95	170	25 * 30	Middle upper lips
R 12	110	150	40 * 30	Right corner mouth
R 13	95	130	25 * 30	Middle lower lips

R 14	70	135	25 *25	Lower left lips
R 15	120	135	25 *25	Lower right lips
R 16	120	160	25 * 25	Upper right lips
R 17	70	160	25 * 25	Upper left lips
R 18	80	180	40 *20	Chin
R 19	25	45	25 * 25	Outer corner of left eyebrows
R 20	70	45	25 * 25	Inner corner of left eyebrows
R 21	100	45	25 * 25	Inner corner of right eyebrows
R 22	145	45	25 * 25	Outer corner of right eyebrows

As an example of allocating the front head area (R1), first, we allocate point (50, 45) from the resized face and rectangle area of with 100 pixels width and 30 pixels height. These specifications allow us to measure the flow value in the front head area as shown in Figure A-2 below.

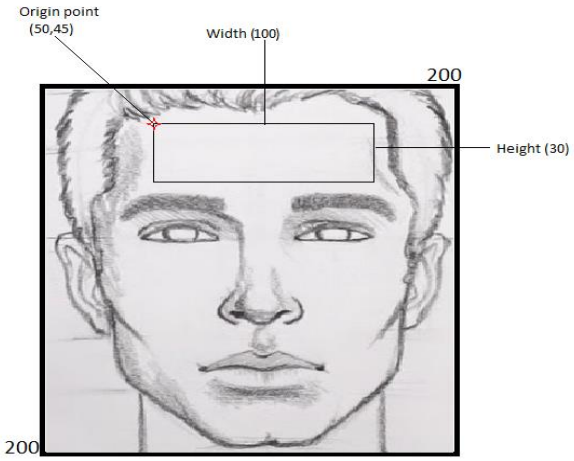


Figure A-2: Allocating the front head area (R1).

Appendix B: The Optical Flow

Optical flow (OPF) can be defined as "*the displacement field for each of the pixels in an image sequence*" or in other words to describe image motion. It is usually applied on image sequences that have a small-time gap between them. In recent work it has been shown that OPF track points or “facial landmarks” in multiple images/video and determine how these points are move. Optical flow algorithm is crafted based on several assumptions [58]:

1) the pixel intensities of an object do not change between consecutive frames.
2) neighbouring pixels have similar motion.

Based on these assumptions, consider we need to track pixel value through image sequence. We need to compute the $I(x, y, t)$, where (I) is the image intensity value at location (x, y) at time (t) which is used to distinguish between reference frame and the current frame. To compute OPF in an image sequence we need to compute the distance between current frame and in the next frame (dt) which will detonate as (dx, dy).

Based on these assumptions consider we need to track pixel ($I(x, y, t)$) where (I) is the image intensity value at location (x, y) at Time (t). The movement in the next frame can be denoted by distance (dx, dy) taken after dt time. Thus, the OPF can be formulated using:

$$I(x, y, t) = I(x + dx\ , y + dy, t + dt) \tag{B.1}$$

After taking Taylor series approximation of right-hand side and removing common terms and dividing by (dt) to get the following equation:

$$fx\,U + fy\,V + ft = 0, \tag{B.2}$$

where $fx = \frac{df}{dx}$; $fy = \frac{df}{dy}$ & $U = \frac{dx}{dt}$; $V = \frac{dy}{dt}$.

The following equation is called the optical flow equation, where (fx) and (fy) are image gradients and (ft) is the gradient along time. However, (U, V) is the unknown and this equation can't be solved with two unknown variables. So, several methods were proposed to solve this problem. Based on the recent work [21, 57] in this Appendix we describe a dense optical flow by Farneback [58].

Two-Frame Motion Estimation Based on Polynomial Expansion by Gunner Farneback

Franeback's approach involves computing all the pixels in an image to identify motion. In their approach they approximate the neighbourhood using polynomial expansion (specifically using quadratic polynomials). The following equation shows the local signal model and expressed in a local coordinate system:

$$f(x) = x^T Ax + b^T x + c, \tag{B.3}$$

where (A) is a symmetric matrix, (b) is a vector and (c) is a scalar. The coefficients are estimated by weighted least square fit to a single pixel values of the neighbourhood.

Based in the hypotheses if polynomial undergoes an ideal translation taking in consideration the exact quadratic polynomial:

$$f_1(x) = x^T A_1 x + b_1^T x + c_1. \tag{B.4}$$

A new signal $f_2(x)$ can be constructed by global displacement d.

$$f_2(x) = x^T A_2 x + b_2^T x + c_2, \quad \text{(B.5)}$$

$$f_2(x) = f_1(x-d) = (x-d)^T A_1 (x-d) + b_1^T (x-d) + c_1, \quad \text{(B.6)}$$

$$= x^T A_1 x + (b_1 - 2A_1 d)^T x + d^T A_1 d - b_1^T d + c_1, \quad \text{(B.7)}$$

$$= x^T A_2 x + b_2^T + c_1, \quad \text{(B.8)}$$

where (based on the brightness constancy assumption):

$$A_2 = A_1, \quad \text{(B.9)}$$

$$b_2 = b_1 - 2A_1 d, \quad \text{(B.10)}$$

The value of the displacement can be computed through:

$$2A_1 d = -(b_2 - b_1), \quad \text{(B.11)}$$

$$d = -\frac{1}{2} A_1^{-1} (b_2 - b_1), \quad \text{(B.12)}$$

where (A) is non-singular. To estimate a displacement value (d) they extract neighbourhood pixels of point (x, y) and point ($x + dx, y + dy$). The polynomial basis is extracted and the computation that is shown above is performed.

Practical Considerations

In their work they assume the entire signal (motion) is a single polynomial and a global translation relating the two signals which are unrealistic. Furthermore, when the assumptions are relaxed, and errors are introduced which can be computed in Equation (B.11). They start by replacing global polynomial in equation (B.3) with local polynomial approximations.

Given two images (A_2, A_1) with expansion coefficients $(A_1(x), b_1(x), c_1(x))$ and $(A_2(x), b_2(x), c_2(x))$ for both images respectively, since $A_2 = A_1$ and according to Equation (9) in practice we need an approximation where,

$$A(x) = \frac{A_1(x)+A_2(x)}{2}, \quad \text{(B.13)}$$

and

$$\Delta b(x) = -\frac{1}{2}A_1^{-1}\big(b_2(x) - b_1(x)\big), \quad \text{(B.14)}$$

Computing primary constraint:

$$A(x)d(x) = \Delta b(x). \quad \text{(B.15)}$$

Equation (B.15) can be solved point wise, but the result can be noisy. On the other hand, they make the assumption that there is small variation in the displacement field. This assumption integrates the information from neighbourhood pixels. So, to satisfy equation (B.15) taking in consideration neighbourhood (I) of (x), or more formally minimizing,

$$\sum_{\Delta x \in I} w(\Delta x)\|A(x+\Delta x)d(x) - \Delta b(x+\Delta x)\|^2. \quad \text{(B.16)}$$

where $(w(\Delta x))$ is the weight function for the neighbourhood points, the minimum value obtained using the following equation for,

$$d(x) = (\sum w\, A^T A)^{-1} \sum w\, A^T\, \Delta b. \quad \text{(B.17)}$$

After dropping some indexing to make the expression more readable the minimum value can be computed using,

$$e(x) = (\sum w\, \Delta b^T\, \Delta b) - d(x)^T \sum w\, A^T \Delta b. \quad \text{(B.18)}$$

Thus, these equations used to compute the flow for every successive frame and to estimate the displacement vector for the entire image sequence or video.

Appendix C: Viola Jones Algorithm

Viola-jones algorithm consist of the following 4 main parts:

1. Haar-like Features

The Haar-like features or rectangle filters have been widely used in object detection and they are named after Haar wavelets. Figure C-1 shows some examples of Haar features (A and B show two rectangular features, C shows three rectangular features and D shows four rectangular features). The Viola-Jones algorithm uses just two-rectangular features.

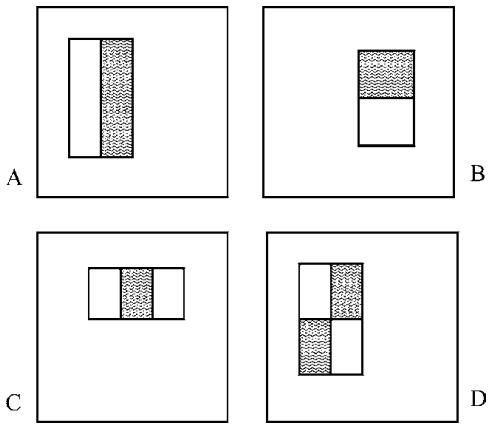


Figure C-1 : Haar-like features [24].

In order to define the subsections of an image, Haar-like features define a detection window that considers adjacent rectangular regions at a specific location, calculating the sum of pixel intensity of each region. It then computes the difference between these sums that defines the image subsection. In order to understand how to choose Haar-like features in the Viola-Jones algorithm, they use some properties common in human faces, for example:

- the eye region is darker than the upper cheeks,
- the nose bridge region is brighter than the eyes.

To reduce the computational time needed to compute the Haar-like features, the Viola-Jones algorithm uses an “integral image”.

2. Integral Image for Rapid Feature Detection

Integral image computes a value at each pixel (x, y) that is the sum of the pixel values above and to the left of (x, y) inclusive as described in [24], where,

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (3.1)$$

where $ii(x, y)$ is the integral image and $i(x', y')$ is the original image.

3. AdaBoost Machine Learning Method

Computing integral images and applying Haar-like features, which are computationally heavy, creates the number of features in a window of size 24*24 applying a five-rectangle feature (2 x two-rectangle features, 2 x three-rectangle features and 1 x four-rectangle) which equals 162,336 [24]. In order to reduce the number of features the Viola-Jones algorithm uses the AdaBoost machine-learning method. A common observation among all faces is that the region of the eyes is darker than the region of the cheeks. Therefore, AdaBoost chooses Haar-like features for face detection as a set of two adjacent rectangles that lie above the eye and the cheek region, as shown in Figure C-2.

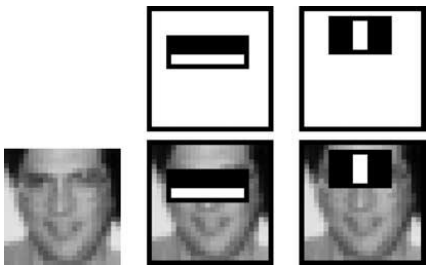


Figure C-2 : Haar-like features defined by AdaBoost [24].

Another Haar-like feature defined by AdaBoost, as shown in Figure C-2, is that the nose region is brighter than the region of the eyes. Therefore, AdaBoost chooses Haar-like features for face detection as a set of three adjacent rectangles that lie above the nose and eye regions. The Viola-Jones algorithm claims that through using AdaBoost, they reduced the number of features from 162,336 to 6,000 [24].

4. Cascaded Classifier to Combine Many Features Efficiently

Computing 6,000 features in a 24*24 window is computationally heavy. By using a cascaded classifier, the main idea is to reject many non-face sub windows without the need to compute 6,000 features, which makes the Viola-Jones algorithm applicable for real-time application

Appendix D: The Machine Learning Algorithm for the Facial Expression Biometric

In this appendix, we use machine-learning techniques to test the emotional biometric. We use a MUG dataset which contains 20 subjects where each subject has on average of three smiles. We use two machine-learning techniques which include: principle component analysis (PCA) and deep convolutional neural network.

1. Principle Component Analysis (PCA)

Principal component analysis (PCA) is a technique that is used for data compression and classification [119]. PCA's original use was to reduce the dimensionality of a dataset. This is done by identifying a new set of linearly uncorrelated variables that contains most significant information about the data. This set of variables is called the principal components (PCs).

PCA algorithms have been applied in a wide-ranging field of studies including face detection, face recognition and gender classification. Furthermore, they have been used as a tool for data analysis and making predictive models.

PCA is applied efficiently in face recognition and gains a high classification rate. Basically, PCA represents the face using eigenfaces, which transform the face into a set of principal components (eigenfaces). PCA uses these eigenfaces to initiate the learning phase. Recognition is done by projecting a new face into the eigenface subspace and using a distance algorithm to identify the appropriate class.

In terms of our research, we apply PCA to the MUG dataset and try to classify it using the smile expression video. The MUG dataset contains 20 subjects expressing smiles with an average of three smiles per subject. For training PCA we use 70% of the data (7,056 images). For testing, we use 30% of the data (one smile video per subject). For classification, we use Euclidean distance to measure a new smile in the smile subspace. The results show a detection rate with 99% using all the 7056-eigenfaces for getting the highest detection rate. This high classification rate is due to a large number of features that have been applied in the PCA training phase which shows that the smile does not affect the face recognition.

2. Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNN) have been a hotspot in computer vision in recent years as they show improvement for the state-of-the-art application. CNN is a deep artificial neural network that is used primarily to classify images, find similarities and accomplish object recognition. Furthermore, these algorithms can identify different objects such as faces, individuals, street signs, tumours and other parts of visual data. According to [171], CNN is inspired by the animal visual cortex which shows that the connectivity pattern between neurons resembles the organisation of the cortex.

In terms of computing, CNN consists of several layers that process and transform data through layers to produce an output. Usually, CNN is trained using a huge number of images in order to perform classification, object detection, segmentation and image processing. According to [172], CNN contains three

main concepts: local receptive fields, shared weights and biases, and activation and pooling.

Local receptive fields in CNN consist of input layer neurons connected to neurons in the hidden layer, where they can be used to create a feature map from the input layer to the hidden layer neurons. For efficient application of this process they apply convolution.

Secondly, in a typical neural network, a CNN has neurons with weights and biases. Where these values are computed using training images, the CNN model learns and updates them with each new training data. On the other hand, in the case of CNNs, the weights and biases are the same for all the hidden neurons in a given layer which shows that the same layers are detecting the same feature. Finally, activation and pooling are used to transform the output of each neuron by using an activation function. One of the well-known activation functions is the Rectified Linear Unit (ReLU) which uses the highest value to map the output of the neuron. Pooling is used to reduce the diminution of the features map produced by the activation function which reduces the number of parameters that the model needs to identify to learn. CNN can have tens or hundreds of hidden layers that learn to detect different features of an image using these three concepts.

Typically, CNN has three ways to analyse images: train a new CNN, transfer learning and use a pre-trained CNN to extract features. Train a new CNN, has a higher accurate output but needs a huge number of images and significant computational resources. Transfer learning uses the knowledge from a specific model and reuses it as the starting point for another model with different data and object. Comparing the computational resources and the data needed to train a

new CNN shows fewer data and fewer computational resources. Finally, where using a pre-trained model to extract features from images, then using some classification algorithms to classify these features.

In our research, we use a pre-trained model which is the VGG model to extract fractures from our dataset. VGG-face is a CNN pre-trained model presented by [173] where they use 2.6 million images to train their model. In our experiment, we use the MUG dataset where we use 20 subjects with an average of three smiles, for training we use 70% of the data which contain 7,056 images, for testing we use the other 30% which contain 2,883 images. The results show a 99.3% correct classification with 0 error rate using layer 34 of the CNN.